



US009249201B2

(12) **United States Patent**  
**Enenkel**

(10) **Patent No.:** **US 9,249,201 B2**  
(45) **Date of Patent:** **\*Feb. 2, 2016**

(54) **HETEROLOGOUS INTRON WITHIN A SIGNAL PEPTIDE**

(71) Applicant: **Barbara Enenkel**, Warthausen (DE)

(72) Inventor: **Barbara Enenkel**, Warthausen (DE)

(73) Assignee: **Boehringer Ingelheim International GmbH**, Ingelheim am Rhein (DE)

(\*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.  
This patent is subject to a terminal disclaimer.

(21) Appl. No.: **14/134,169**

(22) Filed: **Dec. 19, 2013**

(65) **Prior Publication Data**

US 2014/0248660 A1 Sep. 4, 2014

(30) **Foreign Application Priority Data**

Dec. 31, 2012 (EP) ..... 12199813

(51) **Int. Cl.**

**C12P 21/02** (2006.01)  
**C07K 14/47** (2006.01)  
**C12N 15/62** (2006.01)  
**C12N 15/63** (2006.01)  
**C12N 15/67** (2006.01)  
**C12N 15/85** (2006.01)  
**C07K 7/08** (2006.01)  
**C07K 16/00** (2006.01)

(52) **U.S. Cl.**

CPC . **C07K 14/47** (2013.01); **C07K 7/08** (2013.01);  
**C07K 16/00** (2013.01); **C12N 15/625**  
(2013.01); **C12N 15/63** (2013.01); **C12N 15/67**  
(2013.01); **C12N 15/85** (2013.01); **C07K**  
**2317/14** (2013.01); **C07K 2317/52** (2013.01);  
**C07K 2319/00** (2013.01); **C12N 2830/42**  
(2013.01); **C12N 2840/105** (2013.01)

(58) **Field of Classification Search**

None  
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

6,010,884 A 1/2000 Griffiths et al.  
6,972,324 B2 12/2005 Adolf et al.

7,318,923 B2 1/2008 Tsurushita et al.  
7,320,790 B2 1/2008 Hinton et al.  
2005/0090648 A1 4/2005 Tsurushita et al.  
2008/0160048 A1 7/2008 Fuller  
2014/0234893 A1 8/2014 Enenkel  
2014/0242637 A1 8/2014 Enenkel  
2014/0248660 A1 9/2014 Enenkel  
2014/0248665 A1 9/2014 Enenkel

#### FOREIGN PATENT DOCUMENTS

EP 1283263 A1 2/2003  
WO 2002088307 A2 11/2002  
WO 2006071804 A2 7/2006  
WO 2007045465 A1 4/2007  
WO 2009055656 A2 4/2009  
WO 2010010108 A1 1/2010

#### OTHER PUBLICATIONS

Buchman, Andrew R. et al. "Comparison of Intron-Dependent and Intron-Independent Gene Expression" *Molecular and Cellular Biology*, (1988) vol. 8, No. 10, pp. 4395-4405.  
International Search Report for PCT/EP2013/0077046 mailed on Mar. 31, 2014.  
International Search Report for PCT/EP2013/077024 mailed on Apr. 16, 2014.  
International Search Report for PCT/EP2013/077029 mailed May 16, 2014.  
International Search Report for PCT/EP2013/077042 mailed on Jul. 3, 2014.  
Khamlichi, Ahmed Amine et al. "The Effect of Intron Sequences on Expression Levels of Ig cDNAs" *Gene*, (1994) vol. 150, pp. 387-390.  
Neuberger, Michael S. "Expression and regulation of immunoglobulin heavy chain gene transfected into lymphoid cells" *The EMBO Journal*, (1983) vol. 2, No. 8, pp. 1373-1378.  
Neuberger, Michael S. et al. "The Intron Requirement for Immunoglobulin Gene Expression is Dependent upon the Promoter" *Nucleic Acids Research*, (1988) vol. 16, No. 14, pp. 6713-6724.  
Plantier, Jean-Luc et al. "A Factor VIII Minigene Comprising the Truncated Intron I of Factor IX Highly Improves the In Vitro Production of Factor VIII" *Thromb Haemost* (2001) vol. 86, pp. 596-603.

*Primary Examiner* — Jim Ketter

(74) *Attorney, Agent, or Firm* — Michael P. Morris; Atabak R. Royae

(57) **ABSTRACT**

The invention concerns the field of recombinant gene engineering. It concerns novel introns and compositions comprising such introns as well as a method to improve expression of polypeptides from nucleic acids such as cloned genes with heterologous introns, especially genes encoding antibodies and antibody derived fragments, and the production of various polypeptides in eukaryotic host cells using said novel intron sequences as heterologous introns.

**19 Claims, 9 Drawing Sheets**

Figure 1

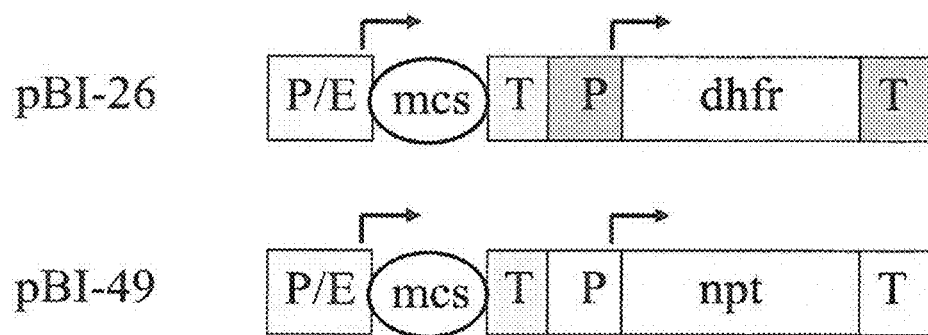


Figure 2

| Recombinant Vector | Description   |
|--------------------|---|
| pBI-26/chIgG1      | mouse/human chimeric IgG1 (genomic DNA with introns within the codon for the first amino acid of the CH1, hinge, CH2 and CH3 domain)  |
| pBI-26/chIgG1-c    | mouse/human chimeric IgG1 (cDNA)  |
| pBI-49/chkappa     | mouse/human chimeric kappa chain (genomic DNA with intron within the codon for the first amino acid of the CL domain)   |
| pBI-49/chkappa-c   | mouse/human chimeric kappa chain (cDNA)   |
| pBI-26/chIgG1B     | mouse/human chimeric IgG1, molecule B (genomic DNA with introns within the codon for the first amino acid of the CH1, hinge, CH2 and CH3 domain)  |
| pBI-26/chIgG1B-c   | mouse/human chimeric IgG1, molecule B (cDNA)  |
| pBI-26/chIgG2B     | mouse/human chimeric IgG2, variable region derived from chIgG1B (genomic DNA with introns within the codon for the first amino acid of the CH1, hinge, CH2 and CH3 domain)  |
| pBI-26/chIgG2B-c   | mouse/human chimeric IgG2, variable region derived from chIgG1B (cDNA)  |
| pBI-26/chIgG4B     | mouse/human chimeric IgG4 with Ser228Pro substitution in hinge region, variable region derived from chIgG1B (genomic DNA with introns within the codon for the first amino acid of the hinge, CH2 and CH3 domain) |
| pBI-26/chIgG4B-c   | mouse/human chimeric IgG4 with Ser228Pro substitution in hinge region, variable region derived from chIgG1B (cDNA)  |
| pBI-49/chkappaB    | mouse/human chimeric kappa chain, molecule B (genomic DNA with intron within the codon for the first amino acid of the CL domain)   |
| pBI-49/chkappaB-c  | mouse/human chimeric kappa chain, molecule B (cDNA)   |
| pBI-26/huIgG1      | humanized IgG1 (genomic DNA with introns within the codon for the first amino acid of the hinge, CH2 and CH3 domain)  |
| pBI-26/huIgG1-c    | humanized IgG1 (cDNA)   |
| pBI-49/hukappa     | humanized kappa chain (genomic DNA with intron within the codon for the first amino acid of the CL domain)  |
| pBI-49/hukappa-c   | humanized kappa chain (cDNA)  |
| pBI-26/hIgG1       | human IgG1 (genomic DNA with introns within the codon for the first amino acid of the hinge, CH2 and CH3 domain)  |
| pBI-26/hIgG1-c     | human IgG1 (cDNA)   |
| pBI-49/hlambda-c   | human lambda chain (cDNA)   |
| pBI-26/Fc-IgG1     | Fc fusion protein with Fc part derived from human IgG1 (genomic DNA with intron within the codon for the first amino acid of the CH3 domain)  |
| pBI-26/Fc-IgG1KO   | Fc fusion protein with Fc part derived from human IgG1 containing Leu234Ala, Leu235Ala substitutions in the CH2 domain (genomic DNA with intron within the codon for the first amino acid of the CH3 domain)      |
| pBI-26/Fc-IgG1-c   | Fc fusion protein with Fc part derived from human IgG1 (cDNA)   |
| pBI-26/Fc-IgG1KO-c | Fc fusion protein with Fc part derived from human IgG1 containing Leu234Ala, Leu235Ala substitutions in the lower hinge region (cDNA)   |

Figure 3

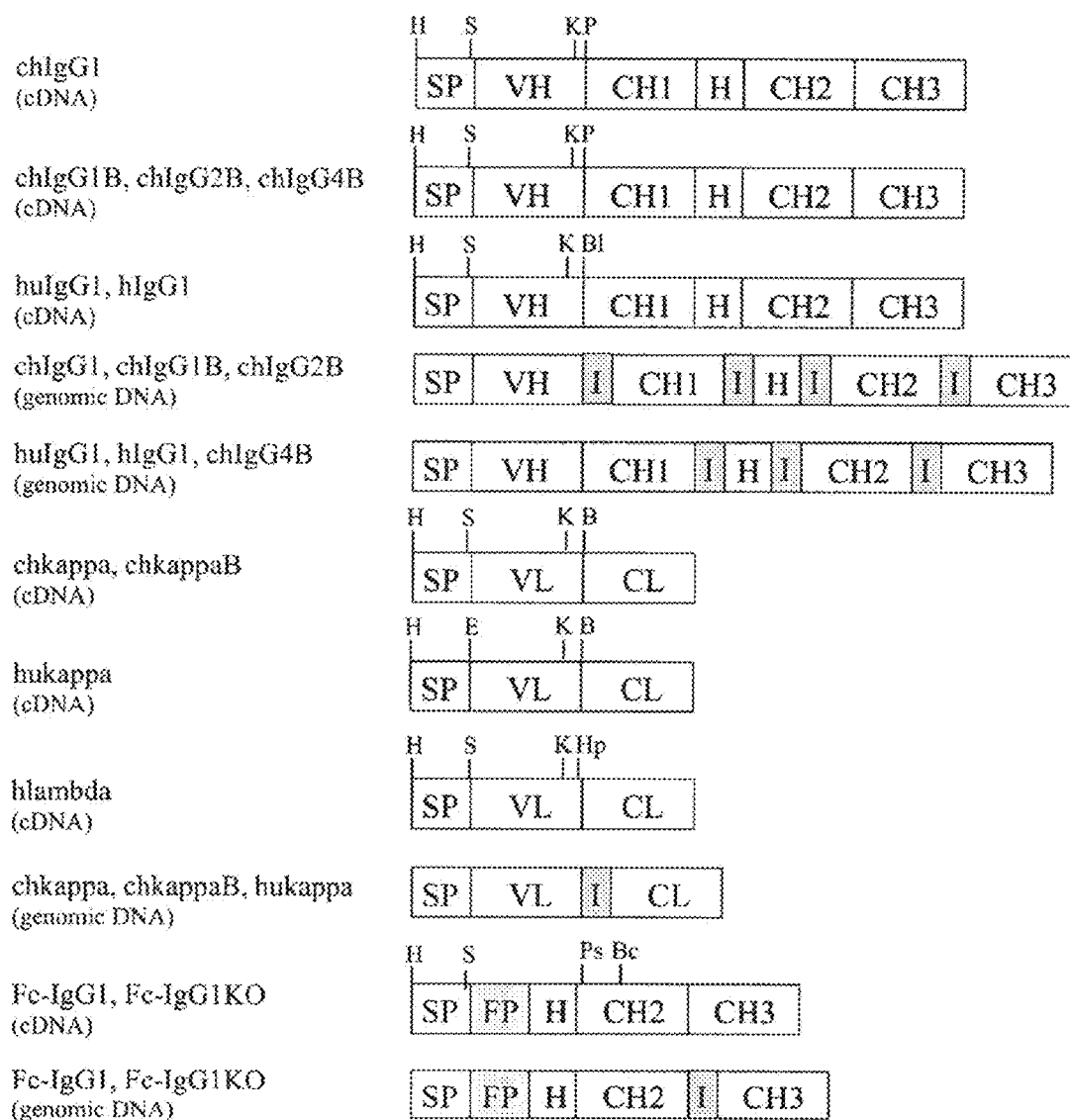


Figure 4

A)

EgIII  
~~~~~

```
1  GTAAAGTGCACCTTTCCTAATAGATCTAATTCTAAACTCTGAGGGGGTTCGGA
51  TGACGTGGCCATTCTTTGCCCTAAAGCATTGAGTTTACTGCAAGGTCAGAA
101 AAGCATGCAAAGCCCTCAGAATGGCTGCAAGAGCTCCAACAAAACAATT
151 TAGAACTTTATTAAAGGAATAGGGGGAAGCTAGGAAGAAACTCAAAACATC
201 AAGATTTTAAATACGCTTCTTGGTCTCCTTGCTATAATTATCTGGGATAA
251 GCATGCTGTTTTCTGTCTGTCCCTAACATGCCCTGTGATTATCCGCAAAC
301 AACACACCCAAGGGCAGAACTTTGTTATACTAACACCATCCTGTTTGCTT
351 CTTTCCTCAG
```

(SEQ ID NO: 1)

B)

```
1  GTAAGTGGCTGGATTGGGTAGGGAAACCGAGGCGGTTCCGCTGAATCGGGT
51  CGAGCACTTGGCGGAGACGCGCGGGCCAACTACTTAGGGACAGTCATGAG
101 GGGTAGGCCCCGCGGCTGCTGCCCTTGCCCATGCCCCGCGGTGATCCCCAT
151 GCTGTGCCAGCCTTTGCCAGAGGCGCTCTAGCTGGGAGCAAAGTCCGGT
201 CACTGGGCAGCACCACCCCCCGGACTTGCAATGGGTAGCCGCTGAGATGGA
251 GCCTGAGCACACGTGACAGGGTCCCTGTTAACGCAGTGTTCCTCTAACTT
301 TCAG
```

(SEQ ID NO: 2)

C)

EgIII  
~~~~~

```
1  GTAAAGTGGCTGGATTGGGTAGATCTGGAAACCGAGGCGGTTCCGCTGAATC
51  GGGTCGAGCACTTGGCGGAGACGCGCGGGCCAACTACTTAGGGACAGTCA
101 TGAGGGGTAGGCCCCGCGGCTGCTGCCCTTGCCCATGCCCCGCGGTGATCC
151 CCATGCTGTGCCAGCCTTTGCCAGAGGCGCTCTAGCTGGGAGCAAAGTC
201 CGGTCACTGGGCAGCACCACCCCCCGGACTTGCAATGGGTAGCCGCTGAGA
251 TGGAGCCTGAGCACACGTGACAGGGTCCCTGTTAACGCAGTGTTCCTCTC
301 CCTTTCAG
```

(SEQ ID NO: 3)

Figure 5

| Recombinant Vector  | Description  |
|---------------------|--|
| pBI-26/chIgG1-g1    | mouse/human chimeric IgG1 with modified human kappa intron (SEQ ID NO: 1) within signal peptide  |
| pBI-49/chkappa-g1   | mouse/human chimeric kappa chain with modified human kappa intron (SEQ ID NO: 1) within signal peptide   |
| pBI-26/chIgG1B-g1   | mouse/human chimeric IgG1, molecule B, with modified human kappa intron (SEQ ID NO: 1) within signal peptide   |
| pBI-26/chIgG2B-g1   | mouse/human chimeric IgG2, variable region derived from chIgG1B, with modified human kappa intron (SEQ ID NO: 1) within signal peptide   |
| pBI-26/chIgG4B-g1   | mouse/human chimeric IgG4 with Ser228Pro substitution in hinge region, variable region derived from chIgG1B, with modified human kappa intron (SEQ ID NO: 1) within signal peptide           |
| pBI-49/chkappaB-g1  | mouse/human chimeric kappa chain, molecule B with modified human kappa intron (SEQ ID NO: 1) within signal peptide   |
| pBI-26/huIgG1-g1    | humanized IgG1 with modified human kappa intron (SEQ ID NO: 1) within signal peptide   |
| pBI-49/hukappa-g1   | humanized kappa chain with modified human kappa intron (SEQ ID NO: 1) within signal peptide  |
| pBI-26/hIgG1-g1     | human IgG1 with modified human kappa intron (SEQ ID NO: 1) within signal peptide   |
| pBI-49/hlambda-g1   | human lambda chain with modified human kappa intron (SEQ ID NO: 1) within signal peptide   |
| pBI-26/Fc-IgG1-g1   | Fc fusion protein with Fc part derived from human IgG1 with modified human kappa intron (SEQ ID NO: 1) within signal peptide   |
| pBI-26/Fc-IgG1KO-g1 | Fc fusion protein with Fc part derived from human IgG1 containing Leu234Ala, Leu235Ala substitutions in the CH2 domain with modified human kappa intron (SEQ ID NO: 1) within signal peptide |
| pBI-26/chIgG1-g4    | mouse/human chimeric IgG1 with hamster dhfr intron (SEQ ID NO: 2) within signal peptide  |
| pBI-49/chkappa-g4   | mouse/human chimeric kappa chain with hamster dhfr intron (SEQ ID NO: 2) within signal peptide   |
| pBI-26/chIgG1B-g4   | mouse/human chimeric IgG1, molecule B with hamster dhfr intron (SEQ ID NO: 2) within signal peptide  |
| pBI-26/chIgG2B-g4   | mouse/human chimeric IgG2, variable region derived from chIgG1B, with hamster dhfr intron (SEQ ID NO: 2) within signal peptide   |
| pBI-26/chIgG4B-g4   | mouse/human chimeric IgG4 with Ser228Pro substitution in hinge region, variable region derived from chIgG1B, with hamster dhfr intron (SEQ ID NO: 2) within signal peptide                   |
| pBI-49/chkappaB-g4  | mouse/human chimeric kappa chain, molecule B with hamster dhfr intron (SEQ ID NO: 2) within signal peptide   |
| pBI-26/huIgG1-g4    | humanized IgG1 with hamster dhfr intron (SEQ ID NO: 2) within signal peptide   |
| pBI-49/hukappa-g4   | humanized kappa chain with hamster dhfr intron (SEQ ID NO: 2) within signal peptide  |
| pBI-26/hIgG1-g4     | human IgG1 with with hamster dhfr intron (SEQ ID NO: 2) within signal peptide  |
| pBI-49/hlambda-g4   | human lambda chain with with hamster dhfr intron (SEQ ID NO: 2) within signal peptide  |
| pBI-26/Fc-IgG1-g4   | Fc fusion protein with Fc part derived from human IgG1 with hamster dhfr intron (SEQ ID NO: 2) within signal peptide   |
| pBI-26/Fc-IgG1KO-g4 | Fc fusion protein with Fc part derived from human IgG1 containing Leu234Ala, Leu235Ala substitutions in the CH2 domain with hamster dhfr intron (SEQ ID NO: 2) within signal peptide         |

Figure 5 (continued)

| Recombinant Vector  | Description   |
|---------------------|---|
| pBI-26/chIgG1-g6    | mouse/human chimeric IgG1 with modified hamster dhfr intron (SEQ ID NO: 3) within signal peptide  |
| pBI-49/chkappa-g6   | mouse/human chimeric kappa chain with modified hamster dhfr intron (SEQ ID NO: 3) within signal peptide   |
| pBI-26/chIgG1B-g6   | mouse/human chimeric IgG1, molecule B with modified hamster dhfr intron (SEQ ID NO: 3) within signal peptide  |
| pBI-26/chIgG2B-g6   | mouse/human chimeric IgG2, variable region derived from chIgG1B, with modified hamster dhfr intron (SEQ ID NO: 3) within signal peptide   |
| pBI-26/chIgG4B-g6   | mouse/human chimeric IgG4 with Ser228Pro substitution in hinge region, variable region derived from chIgG1B, with modified hamster dhfr intron (SEQ ID NO: 3) within signal peptide           |
| pBI-49/chkappaB-g6  | mouse/human chimeric kappa chain, molecule B with modified hamster dhfr intron (SEQ ID NO: 3) within signal peptide   |
| pBI-26/huIgG1-g6    | humanized IgG1 with modified hamster dhfr intron (SEQ ID NO: 3) within signal peptide   |
| pBI-49/hukappa-g6   | humanized kappa chain with modified hamster dhfr intron (SEQ ID NO: 3) within signal peptide  |
| pBI-26/hIgG1-g6     | human IgG1 with with modified hamster dhfr intron (SEQ ID NO: 3) within signal peptide  |
| pBI-49/hlambda-g6   | human lambda chain with with modified hamster dhfr intron (SEQ ID NO: 3) within signal peptide  |
| pBI-26/Fc-IgG1-g6   | Fc fusion protein with Fc part derived from human IgG1 with modified hamster dhfr intron (SEQ ID NO: 3) within signal peptide   |
| pBI-26/Fc-IgG1KO-g6 | Fc fusion protein with Fc part derived from human IgG1 containing Leu234Ala, Leu235Ala substitutions in the CH2 domain with modified hamster dhfr intron (SEQ ID NO: 3) within signal peptide |

Figure 6

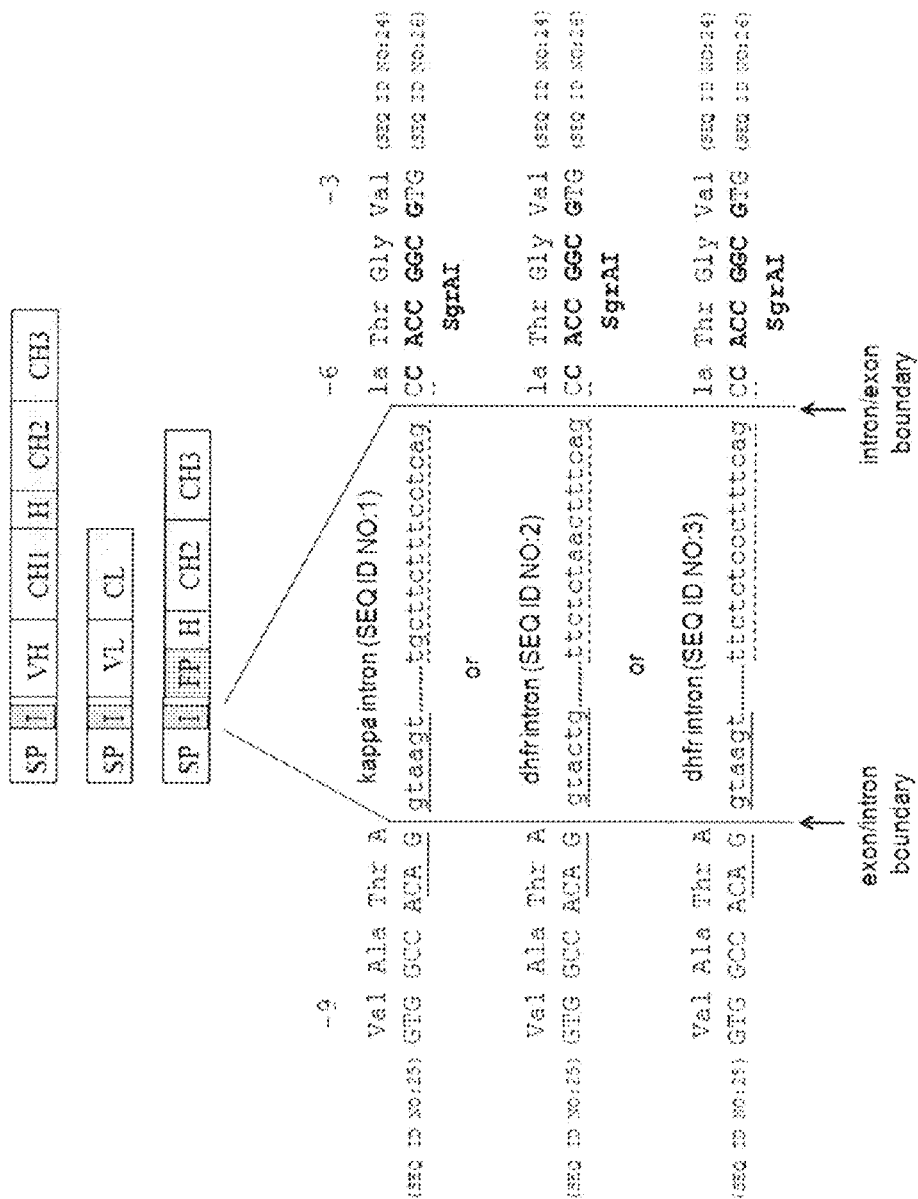




Figure 7

| Antibody | Vector 1          | Vector 2           | % Expression of cDNA<br>(cDNA = 100 %) |
|----------|-------------------|--------------------|--|
| chIgG1   | pBI-26/chIgG1-c   | pBI-49/chkappa-c   | 100                                    |
|          | pBI-26/chIgG1     | pBI-49/chkappa     | 237                                    |
|          | pBI-26/chIgG1-g1  | pBI-49/chkappa-g1  | 367                                    |
| chIgG1B  | pBI-26/chIgG1B-c  | pBI-49/chkappaB-c  | 100                                    |
|          | pBI-26/chIgG1B    | pBI-49/chkappaB    | 186                                    |
|          | pBI-26/chIgG1B-g1 | pBI-49/chkappaB-g1 | 384                                    |
| chIgG1B  | pBI-26/chIgG1B-c  | pBI-49/chkappaB-c  | 100                                    |
|          | pBI-26/chIgG1B    | pBI-49/chkappaB    | 212                                    |
|          | pBI-26/chIgG1B-g4 | pBI-49/chkappaB-g4 | 201                                    |
| hulgG1   | pBI-26/hulgG1-c   | pBI-49/hukappa-c   | 100                                    |
|          | pBI-26/hulgG1     | pBI-49/hukappa     | 270                                    |
|          | pBI-26/hulgG1-g1  | pBI-49/hukappa-g1  | 351                                    |
| hIgG1    | pBI-26/hIgG1-c    | pBI-49/hlambd-c    | 100                                    |
|          | pBI-26/hIgG1      | pBI-49/hlambd-c*   | 96                                     |
|          | pBI-26/hIgG1-g1   | pBI-49/hlambd-g1   | 169                                    |
| chIgG2B  | pBI-26/chIgG2B-c  | pBI-49/chkappaB-c  | 100                                    |
|          | pBI-26/chIgG2B    | pBI-49/chkappaB    | 240                                    |
|          | pBI-26/chIgG2B-g1 | pBI-49/chkappaB-g1 | 298                                    |
| chIgG4B  | pBI-26/chIgG4B-c  | pBI-49/chkappaB-c  | 100                                    |
|          | pBI-26/chIgG4B    | pBI-49/chkappaB    | 252                                    |
|          | pBI-26/chIgG4B-g1 | pBI-49/chkappaB-g1 | 427                                    |

\*cDNA variant used

Figure 8

| SEQ ID NO: | Signal Peptide Sequence and Position of Intron   |
|------------|--|
| 4          | MetGlyTrpSerCysIleIleLeuPheLeuValAlaThrAlaThrGlyValHisSer<br>ACAG:CN                             |
| 5          | MetGluThrAspThrLeuLeuLeuTrpValLeuLeuLeuTrpValProGlySerSerGly<br>GGHAG:Y                          |
| 6          | MetGlyTrpSerTrpIlePheLeuPheLeuLeuSerGlyThrAlaGlyValLeuSer<br>ACAG:CN                             |
| 7          | MetAspTrpThrTrpArgValPheCysLeuLeuAlaValAlaProGlyAlaHisSer<br>GTAG:CN                             |
| 8          | MetAspMetArgValProAlaGlnLeuLeuGlyLeuLeuLeuLeuTrpPheProGlyAlaArgCys<br>CAG:TTR<br>CAG:CTN         |
| 9          | MetAspMetArgValProAlaGlnLeuLeuGlyLeuLeuLeuLeuTrpLeuProGlyAlaArgCys<br>CAG:TTR<br>CAG:CTN         |
| 10         | MetGluPheGlyLeuSerTrpLeuPheLeuValAlaIleLeuLysGlyValGlnCys<br>GTAG:CN                             |
| 11         | MetGluAlaProAlaGlnLeuLeuPheLeuLeuLeuLeuTrpLeuProAspThrThrGly<br>CCAG:CN                          |
| 12         | MetAlaValLeuGlyLeuLeuPheCysLeuValThrPheProSerCysValLeuSer<br>CCHAG:Y                             |
| 13         | MetValSerThrProGlnPheLeuValPheLeuLeuPheTrpIleProAlaSerArgSer<br>SCMAG:Y                          |
| 14         | MetGluThrAspThrIleLeuLeuTrpValLeuLeuLeuTrpValProGlySerThrGly<br>GGHAG:R                          |
| 15         | MetAsnPheGlySerSerLeuIlePheLeuValLeuValLeuLysGlyValGlnCys<br>TCAAG:Y<br>AGYAG:Y                  |
| 16         | MetGlyTrpSerCysIlePheLeuPheLeuValAlaThrAlaThrGlyValHisSer<br>ACAG:CN                             |
| 17         | MetAlaProSerSerProArgProAlaLeuProAlaLeuLeuValLeuLeuGlyAlaLeuPheProGlyProGly<br>CCAG:CN<br>AsnAla |
| 18         | MetLysValSerAlaAlaLeuLeuCysLeuLeuLeuIleAlaAlaThrPheIleProGlnGlyLeuAla<br>GCAG:CN                 |
| 19         | MetGlyValLeuLeuThrGlnArgThrLeuLeuSerLeuValLeuAlaLeuLeuPheProSerMetAlaSerMetAla<br>CCMAG:Y        |
| 20         | MetGlyValArgHisProProCysSerHisArgLeuLeuAlaValCysAlaLeuValSerLeuAlaThrAla<br>GTHAG:Y              |
| 21         | MetTrpGlnLeuLeuLeuProThrAlaLeuLeuLeuLeuValSerAlaGlyMetArgThr<br>TCAG:CN                          |
| 22         | MetLeuGlyProCysMetLeuLeuLeuLeuLeuLeuLeuGlyLeuArgLeuGlnLeuSerLeuGly<br>CTCAG:Y<br>YTAAG:Y         |
| 23         | MetAlaProValAlaValTrpAlaAlaLeuAlaValGlyLeuGlnLeuTrpAlaAlaAlaHisAla<br>GCAG:CN                    |

colon: site of intron insertion, N = any base, Y = C or T, R = A or G, H = C or A or T, M = C or A

1

## HETEROLOGOUS INTRON WITHIN A SIGNAL PEPTIDE

### SEQUENCE LISTING

The instant application contains a Sequence Listing which has been submitted electronically in ASCII format and is hereby incorporated by reference in its entirety. Said ASCII copy, created on Feb. 17, 2014, is named 01-2885\_SL.txt and is 10,306 bytes in size.

### BACKGROUND OF THE INVENTION

#### 1. Technical Field

The invention concerns the field of recombinant gene engineering. It concerns novel introns and compositions comprising heterologous introns as well as a method to improve expression of polypeptides from nucleic acids such as cloned genes, especially genes encoding antibodies and antibody-derived fragments, and the production of various polypeptides in eukaryotic host cells using said novel intron sequences.

#### 2. Background

The market for biopharmaceuticals for use in human therapy continues to grow at a high rate with more than 300 biopharmaceuticals already approved, many more in clinical development and estimated sales of more than 167 billions by 2015. Currently, an increasing number of biopharmaceuticals is produced from mammalian cells due to their ability to correctly process and modify human proteins. Therefore the recombinant proteins are compatible with humans both functionally and pharmacokinetically. A shortcoming compared to prokaryotic expression systems is often the significantly lower protein expression level resulting in higher drug costs. Successful and high yield production of biopharmaceuticals from mammalian cells is thus crucial and is governed by various factors including host cell line, expression system, gene copy number, cell growth and productivity, secretion efficiency of the protein, culture and feed media, production and purification process, protein structure and sequence, protein stability and formulation. Expression of the recombinant protein requires an expression vector encoding the desired gene of interest. Several methods have been employed to optimize expression vectors for efficient protein production. Gene expression is regulated on transcriptional and translational levels. Hence many methods pertain to the identification and optimization of strong promoters and enhancers to improve the efficiency with which protein encoding genes are transcribed. Examples of these are the CMV immediate early promoter and enhancer, SV40 promoter and enhancer, elongation factor (EF) promoter, Polyoma enhancer, and chicken [beta]-actin promoter. Likewise, strong polyadenylation signal sequences such as bovine growth hormone (BGH) and SV40 polyadenylation sites that stabilize mRNAs and enhance transcription termination are also used to augment the protein expression from genes encoded by the expression vectors. Among the methods to improve the efficiency with which the resultant mRNA is translated are the use of translation initiation sites (AUG), optimal ribosome binding sites such as the Kozak sequence or internal ribosome entry sites (IRES) and the tripartite leader element (TPL) from adenovirus.

Another common approach to improve expression is to increase the gene copy number. This can be achieved by transfecting cells with selectable, amplifiable marker genes such as dihydrofolate reductase (DHFR) or glutamine synthetase (GS) genes and growing the cells in the presence of

2

selective agents such as methotrexate in case of DHFR or methionine sulfoximine in case of GS.

By the chance integration of the expression vectors in the host cell genome, cells are obtained with different levels of expression of the desired gene product, as its expression is not determined solely by the strength of the transcriptional and translational regulatory elements described above. The chromatin structure present at the integration site can affect the level of expression both negatively and positively. Increasingly, therefore, cis-active elements which positively influence the expression at the chromatin level are integrated in expression vectors. These include locus control regions (LCR), scaffold/matrix attachment regions (S/MARs), ubiquitous chromatin opening elements (UCOE), expression augmenting sequence elements (EASE), transcription or expression enhancing elements (TE element) or stimulatory and anti-repressor elements (STAR).

Even though there exist prior art elements to increase the protein expression by modulating the expression vector, there is further need to identify regulatory elements to further increase the productivity of a recombinant production cell line.

### SUMMARY OF THE INVENTION

It is an objective of this invention to provide mammalian gene expression vectors for driving strong and stable gene expression in eukaryotic cells.

To achieve this objective the present invention provides novel sequences, expression vectors and a method of producing higher levels of recombinant proteins in eukaryotic cells, which pertains to the use and selection of non coding sequence elements such as introns. Introns normally form an integral part of eukaryotic genes as intervening sequences between exons. They are precisely deleted from the primary transcript by a process known as RNA splicing to form mature messenger RNA (mRNA). RNA splicing is one of the main mechanisms of qualitative and quantitative regulation of gene expression in eukaryotic genomes. Alternative splicing is a major mechanism for the enhancement of transcriptome and proteome diversity. By inclusion of different exons in the mRNA multiple transcripts are obtained from the same gene. However, splicing has also an impact on transcription, mRNA export, mRNA stability and even protein translation. Because of the beneficial effect of the intron during post-transcriptional processes, the complementary DNA (cDNA) version of most natural intron-containing genes, for example immunoglobulin genes, is expressed quite poorly in mammalian cells even if strong transcriptional promoters are used for expression. The presence of cryptic splice sites in the gene sequences can lead to alternative splicing events, especially in intron-containing genes, and thus aberrant gene expression. This might lead to unwanted by-products with altered protein sequences and thus properties and functions. A variety of algorithms might be used to identify putative splicing sites in terms of probability of usage but the success varies. The probability of aberrant splicing events can be lowered by using a cDNA version of the gene of interest. cDNAs are also preferred if the genomic versions of the genes are too large in size to incorporate them into useful expression vectors. The downside is frequently a lower expression of the protein of interest. This is routinely overcome by the optimization of the cDNA sequences thereby removing cryptic splice sites, direct repeats, secondary structure elements and other motifs interfering with expression. Also, the GC content for improved RNA stability and the codon usage are routinely optimized. The process of gene optimization is routinely performed on

sequences without introns. In general, it is challenging to improve the level of recombinant gene expression in eukaryotic cells. Frequently, due to the reasons above, introns within the coding regions are not used to increase recombinant gene expression of genes, which encode proteins of interest, but such proteins of interest are routinely expressed as cDNAs having no introns at all.

The present invention solves this problem by providing novel introns and methods of preparing and selecting mammalian cell lines which allow a high expression of heterologous gene products, preferably biopharmaceutically relevant polypeptides or proteins. Surprisingly, it has been found that these newly identified introns outperform the natural genomic DNA set-up as well as the corresponding cDNA set-up with optimized nucleotide sequences and lead to higher productivity of producer cells.

The present invention provides intron sequences that increase the expression of heterologous nucleotide sequences encoding products of interest, especially nucleotide sequences encoding antibodies and antibody-derived fragments, in a eukaryotic host cell. The intron sequences of the present invention have the advantageous technical effect that the expression of transcription units into which these intron sequences are placed is increased (see FIG. 7).

The present invention provides heterologous intron sequences, which are intron sequences placed at a sequence position in an exon encoding at least part of a signal peptide sequence different from the intron position(s) in the native eukaryotic gene or into a eukaryotic, prokaryotic or synthetic gene, i.e. positions which naturally do not contain an intron within a signal peptide sequence. In native immunoglobulin genes of for example mouse or human origin the introns within the signal peptide sequence are located within the codon encoding the amino acid at position -4 (counting backwards from the 3' end of the amino acid sequence of the signal peptide) of the signal peptide sequence of both heavy and light chains between the first and second nucleotide of the codon. In the present invention the heterologous intron is located within the immunoglobulin signal peptide sequence of transcription units at positions other than the native intron positions described above. It is introduced at a new position preferably in a way that a functional and efficient splice donor site and a functional and efficient splice acceptor site is obtained. More precisely, the heterologous intron is inserted within an immunoglobulin exon, whereby the 5' and 3' ends of said exon are defined as occurring in a corresponding native immunoglobulin gene. Alternatively, the heterologous intron is placed within signal peptide sequences of non-immunoglobulin origin.

#### DESCRIPTION OF THE FIGURES

FIG. 1: Basic Expression Vectors

FIG. 1 schematically shows the design of the expression vector pBI-26 and pBI-49 used for the transfection of CHO-DG44 cells. "P/E" means a composite unit that contains both CMV enhancer and promoter element, "P" on its own a promoter element and "T" a termination signal for transcription, which is required for polyadenylation of transcribed messenger RNA. For cloning of the gene of interest a sequence region with multiple cutting sites for restriction endonucleases (multiple cloning sites—"mcs") is inserted after the promoter/enhancer element. For termination of the transcript of the gene of interest the termination signal of the hamster growth hormone is used. The amplifiable selectable marker dihydrofolate reductase is abbreviated to "dhfr" and is under the control of the promoter and termination signal of the hamster

dhfr gene. The selectable marker neomycin phosphotransferase is abbreviated to "npt" and is under the control of the SV40 early promoter and a thymidine kinase termination signal. The position and direction of transcription initiation within each transcription unit is indicated by an arrow.

FIG. 2: Summary of Recombinant Expression Vectors Used as Control

FIG. 2 summarizes the recombinant expression vectors which are used as a control. The genes of interest are antibody heavy chains of IgG1, IgG2 or IgG4 isotype, antibody kappa or lambda light chain and Fc fusion proteins in which the Fc part is derived from a human IgG1. Chimeric antibody heavy or light chain genes in which the variable regions of the heavy or light chains are of mouse origin and the constant regions of human origin are indicated by the prefix "ch". Antibody genes encoding heavy or light chains with humanized variable regions are indicated by the prefix "hu" and completely human antibody heavy or light chain genes by the prefix "h". "KO" means that the constant domain CH2 of human IgG1 origin contains the amino acid substitutions Leu234Ala and Leu235Ala to reduce antibody mediated effector functions. The hinge region of the IgG4 molecule contains the amino acid substitution Ser228Pro to stabilize the intermolecular disulfide bridges in the hinge region and thus reduce the occurrence of half molecules. Numbering of the amino acids in the constant domains and hinge regions is according to the EU index in Kabat et al. (1983), "Sequences of proteins of immunological interest", US Dept. Health and Human Services. Genes are either cloned as cDNA, marked with the suffix "c", or as genomic DNA versions. In the latter the intron and exon sequences of the constant regions are derived from the natural genomic gene sequences. The introns are in their natural positions located always within the codon encoding for the first amino acid of a constant domain or the hinge region.

FIG. 3: Gene Structure of Genes of Interest in Control Vectors

FIG. 3 schematically shows the gene structure of the genes of interest encoded in the recombinant control vectors. The genes are either cloned as cDNA or as genomic DNA versions. In the genomic DNA the introns are in their natural positions located always within the codon encoding for the first amino acid of a constant domain or the hinge region. "VH" means the variable region of an antibody heavy chain and "VL" the variable region of an antibody light chain. The constant domains of the antibody heavy chains of IgG1, IgG2 or IgG4 isotype are abbreviated to "CH1", "CH2" and "CH3" and the hinge region to "H". The constant domains of the antibody kappa or lambda light chains are abbreviated to "CL". The Fc fusion proteins contain the constant domains "CH2" and "CH3" of the antibody heavy chain and part of the hinge region "H". Introns are abbreviated to "I", signal peptide sequences to "SP" and the fusion partner of the Fc fusion proteins to "FP". Restriction enzyme sites for subcloning of intron sequences into new positions are "S" (=SgrAI), "E" (=EcoRV) and "H" (=HindIII). The Fc fusion protein is abbreviated to "Fc". Chimeric antibody genes in which the variable regions of the heavy or light chains are of mouse origin and the constant domains of human origin are indicated by the prefix "ch". Antibody heavy or light chain genes with humanized variable regions are indicated by the prefix "hu" and completely human antibody heavy or light chain genes by the prefix "h". "KO" means that the constant domain CH2 of human IgG1 origin contains the amino acid substitutions Leu234Ala and Leu235Ala (numbering of the amino acids in the constant domains according to the EU index in Kabat et al.

(1991), "Sequences of proteins of immunological interest", US Dept. Health and Human Services.).

#### FIG. 4: Intron Sequences

FIG. 4 shows the nucleotide sequences of the (A) de novo synthesized modified intron of a human antibody kappa light chain (SEQ ID NO: 1), the (B) dihydrofolate reductase intron (SEQ ID NO: 2) derived from the Chinese hamster (*Cricetus griseus*) and the (C) de novo synthesized modified dihydrofolate reductase intron (SEQ ID NO: 3) derived from the Chinese hamster (*Cricetus griseus*). The intron located splice acceptor site is underlined with a broken line, the intron located part of the splice donor site with a solid line and the putative branch site with a double line. The 5' region of the modified kappa intron and the 5' region of the modified dihydrofolate reductase intron contain a single restriction enzyme site for BglII which can be used for cloning purposes. Boxed nucleotides indicate the first triplets which could act as a stop codon in each of the three possible reading frames in case splicing of the intron at the splice donor site does not occur.

#### FIG. 5: Recombinant Expression Vectors Encoding Genes with Introns in Signal Peptide Sequences

FIG. 5 summarizes the recombinant expression vectors encoding genes of interest in which a single intron is placed in nucleotide sequences encoding an immunoglobulin signal peptide. The genes of interest are antibody heavy chains of IgG1, IgG2 or IgG4 isotype, antibody kappa or lambda light chain and Fc fusion proteins in which the Fc part is derived from a human IgG1. Placement of the modified human kappa light chain intron with SEQ ID NO: 1 within the signal peptide sequence is marked with the suffix "g1", placement of the hamster dhfr intron with SEQ ID NO: 2 with the suffix "g4" and placement of the modified hamster dhfr intron with SEQ ID NO: 3 with the suffix "g6". Chimeric antibody genes in which the variable regions of the heavy or light chains are of mouse origin and the constant domains of human origin are indicated by the prefix "ch". Antibody genes encoding heavy or light chains with humanized variable regions are indicated by the prefix "hu" and completely human antibody heavy or light chain genes with the prefix "h". "KO" means that the constant domain CH2 of human IgG1 origin contains the amino acid substitutions Leu234Ala and Leu235Ala. Numbering of the amino acids in the constant domains of the human heavy chains is according to the EU index in Kabat et al. (1991), "Sequences of proteins of immunological interest", US Dept. Health and Human Services.

#### FIG. 6: Position of Heterologous Introns in Signal Peptide Sequences

FIG. 6 shows schematically the gene structure of the heavy and light chain antibody genes in which the intron has been positioned within an immunoglobulin signal peptide sequence. "VH" means the variable region of an antibody heavy chain and "VL" the variable region of an antibody light chain. The constant regions of the antibody heavy chains of IgG1, IgG2 or IgG4 isotype are abbreviated to "CH1", "CH2" and "CH3" and the hinge region to "H". The constant regions of the antibody kappa or lambda light chains are abbreviated to "CL". The introns are abbreviated to "I", signal peptide sequences to "SP" and the fusion partner of the Fc fusion proteins to "FP".

The nucleotide sequences of the exon/heterologous intron and heterologous intron/exon boundaries as well as the immediate nucleotide sequences flanking these boundaries are shown below. Capital letters indicate the nucleotides of the coding region and small letters the non-coding nucleotides within the heterologous intron region. Splice donor sites and splice acceptor sites are underlined with solid lines and broken lines, respectively. The predicted amino acids

(3-letter code) in the coding regions are shown above the coding nucleotide sequence. Numbering of the amino acids indicates their position within the immunoglobulin signal peptide sequences whereby positions are counted backwards from the 3' end of the signal peptide amino acid sequence. Restriction enzyme sites are marked with bold letters and the corresponding restriction enzyme is indicated below. Genes contain either the modified human kappa light chain intron with SEQ ID NO: 1, the hamster dhfr intron with SEQ ID NO: 2 or the modified hamster dhfr intron with SEQ ID NO: 3. FIG. 6 shows nucleotides 1-6 and 346-360 of the "kappa intron (SEQ ID NO:1)", nucleotides 1-6 and 290-304 of the "dhfr intron (SEQ ID NO:2)" and nucleotides 1-6 and 294-308 of the "dhfr intron (SEQ ID NO:3)".

#### FIG. 7: Impact of Signal Peptide Intron on Expression of IGGs

In at least two independent series CHO-DG44 cells are co-transfected (n=3) with expression vectors encoding various IgG1, IgG2 and IgG4 heavy and light chain gene formats. Chimeric antibodies in which the variable regions of heavy and light chain are of mouse origin and the constant regions of human origin are indicated by the prefix "ch". Antibodies with humanized variable regions are indicated by the prefix "hu" and completely human antibodies by the prefix "h". Genes in the expression vectors are either cloned as cDNA, marked with the suffix "c", or as genomic DNA versions (no suffix). In the latter the intron and exon sequences of the constant regions are derived from the natural genomic gene sequences containing the introns in their natural positions located always within the codon encoding for the first amino acid of a constant domain or the hinge region. Non-natural placement of the modified human kappa light chain intron with SEQ ID NO: 1 within the signal peptide sequence is marked with the suffix "g1" whereas placement of the hamster dhfr intron with SEQ ID NO: 2 in the same position is marked with the suffix "g4". After a period of 48-72 hours the antibody titers in the supernatants are determined using a generic IgG ELISA. To correct for transfection efficiency cells are co-transfected with a plasmid encoding secreted alkaline phosphatase (SEAP) as reporter protein and the SEAP activity is measured. The average antibody titers obtained are used to calculate the mean percentage of expression whereby the cDNA setting within each transfection series for a certain antibody of Fc fusion protein is taken as 100%. In general, the highest antibody expression is obtained from the intron-containing gene sequences with the genes containing the heterologous modified human kappa light chain intron or the hamster dhfr intron placed within the signal peptide sequence performing best.

#### FIG. 8: Signal Peptide Sequences and Intron Positions

FIG. 8 shows examples of signal peptide sequences. The amino acid sequence of each signal peptide is indicated in the three letter code. Codons encoding the underlined amino acid pairs constitute suitable sites for the positioning of a heterologous intron. The preferred nucleotide sequences of each codon contributing to an efficient splice donor and splice acceptor site are indicated underneath the underlined amino acids. Sources of the signal peptides are as follows: mouse immunoglobulin heavy chain (SEQ ID NO: 4, 6, 12, 15 and 16), mouse immunoglobulin kappa light chain (SEQ ID NO: 5, 13 and 14), human immunoglobulin heavy chain (SEQ ID NO: 7 and 10), human immunoglobulin kappa light chain (SEQ ID NO: 8, 9 and 11), human intercellular adhesion molecule (SEQ ID NO: 17), human monocyte chemoattractant protein 1 (SEQ ID NO: 18), human oncostatin M (SEQ ID NO:19), human acid alpha glucosidase (SEQ ID NO:20), human Fc gamma receptor III (SEQ ID NO: 21), human

placental alkaline phosphatase (SEQ ID NO: 22) and human tumor necrosis factor receptor (SEQ ID NO:23).

#### DETAILED DESCRIPTION OF THE INVENTION

The impact of the novel introns of the present invention is demonstrated for various proteins of interest, such as immunoglobulin G1, G2, G4 and Fc fusion proteins.

FIG. 7 shows the data of at least 2 independent transient transfection series performed in triplicate. Overall, cells transfected with heterologous intron containing genes containing introns show 1.7-4.3 fold higher expression of mouse/human chimeric, humanized or human IgG1, IgG2 or IgG4 antibodies compared to cells transfected with vectors encoding the cDNAs of the respective antibodies even though the sequences of the variable regions are optimized. Surprisingly, all heterologous intron-containing gene set-ups in which a modified single intron sequence derived from a human kappa gene (SEQ ID NO: 1, FIG. 4A) is placed in new positions within the signal peptide of heavy and light chain genes are outperforming the corresponding natural genomic gene versions. Also, the natural intron derived from the hamster dihydrofolate reductase gene (SEQ ID NO: 2, FIG. 4B) leads as a single heterologous intron to an expression comparable to the corresponding natural genomic gene version with 3 introns if placed in the signal peptide sequence of transcription units encoding heavy and light chain gene. The general embodiments "comprising" or "comprised" encompass the more specific embodiment "consisting of". Furthermore, singular and plural forms are not used in a limiting way.

Terms used in the course of this present invention have the following meaning.

The terms "intron" as used herein, refer to a non-coding or intervening polynucleotide sequence of varying length, normally present within many eukaryotic genes, which is removed from a newly transcribed mRNA precursor by the process of splicing for which highly conserved sequences at or near either end of the intron are necessary. In general, the process of splicing requires that the 5' and 3' ends of the intron be correctly cleaved and the resulting ends of the mRNA be accurately joined, such that a mature mRNA having the proper reading frame for protein synthesis is produced. Many splice donor and splice acceptor sites, meaning the sequences immediately surrounding the exon-intron- and intron-exon-boundaries, have been characterized and described and are known to the skilled artisan. Normally, the sequence of mammalian introns begins with GT and ends with AG with a few minor introns starting with AT and ending with AC. Introns can be positioned within the actual coding region of a gene or in the 5' and/or 3' untranslated region of a gene. Introns are contained within the mRNA precursor (unspliced mRNA). Each intron contains at least three highly conserved sequence elements essential for splicing: a 5' splice site, a branch point and a 3' splice site containing a run of pyrimidines called a polypyrimidine tract. The branch point or branch site is usually located approximately between 10 and 60 nucleotides upstream of the 3' splice site and forms during the splicing process via its conserved adenosine residue a lariat structure with the 5' splice site. The consensus sequence for the branch point is reported as YNYRAY (with Y=T or C, R=A or G, N=any base; the conserved adenosine involved in the lariat formation is underlined) whereby in mammals the preferred branch point sequence is TACTAAC (Zhuang et al., PNAS 86, 2752-2756, 1989). Often the branch point is adjacent to or within the polypyrimidine tract of the splice acceptor site.

"Splicing" or "spliced" refers to the mechanism by which a single functional RNA molecule is produced by the removal of one or more intron sequences during the processing of the primary transcript. Thereby splice donor sites interact with splice acceptor sites to allow splicing of the RNA or transcript and thus excision of the intron(s) bounded by the splice donor and acceptor sites. For each transcript the splice donor site splices with only one splice acceptor site. In case of alternative splicing the splice donor site splices within the pools of transcripts with more than one splice acceptor site leading to a heterogeneous pool of transcripts. "Spliced mRNA" or "spliced transcript" refers herein to mRNA or transcripts produced by either removal of one or more intron sequences or by constructing a cDNA which when transcribed produces an mRNA or transcript having the same properties as an mRNA or transcript which had been subject to splicing but from which no nucleotide sequences had in fact been removed.

The term "splice site" refers to a specific sequence within a polynucleotide sequence that is present at either the 5' end or the 3' end of an intron. Recognition of these sites by the splicing machinery is followed by the excision of an intron or a polynucleotide sequence flanked by these sites. The term splice sites includes naturally occurring, engineered or synthetic, consensus or cryptic splice sites.

The terms "splice donor site", "5' splice site" or "SD" as used herein refer to a conserved sequence of a polynucleotide sequence surrounding the exon-intron boundary at the 5' end of an intron that marks the start of the intron and its boundary with the preceding exon sequence. The consensus sequence for a splice donor site consists of the sequence MAG: GTRAGT (with M=C or A, R=A or G and the colon denoting the site of exon-intron boundary=cleavage site) (Ohshima et al., J. Mol. Biol. 195, 247-259, 1987).

The terms "splice acceptor site", "3' splice site" or "SA" as used herein refer to a conserved sequence of a polynucleotide sequence surrounding the intron-exon boundary at the 3' end of an intron that marks the end of the intron and its boundary with the following exon sequence. The consensus sequence for a splice acceptor site consists of the sequence Y<sub>11-40</sub>NYAG:R (SEQ ID NO: 27) (with Y=pyrimidine C or T, R=A or G, N=any base and the colon denoting the site of intron-exon boundary=cleavage site) (Ohshima et al., J. Mol. Biol. 195, 247-259, 1987).

Splice donor and splice acceptor sites are well known in the art and any may be utilized in the present invention. For a review see Ohshima et al., J. Mol. Biol. 195, 247-259, 1987. These elements can be found, inter alia, in the art or derived from consensus sequences, either empirically by inserting, deleting or substituting nucleotides, or by using software capable of predicting splicing sequences. One preferred splice donor and splice acceptor site in this invention is the consensus splice donor and splice acceptor site mentioned above. However, other sequences with sufficient splicing efficiency can be used as well. Efficient splice donor and acceptor sites suitable for this invention can be readily determined using techniques for measuring the efficiency of splicing. Intron splicing efficiency is readily determined by quantifying the spliced transcripts versus the full-length, unspliced transcripts that contain the intron(s), using methods known in the art such as quantitative PCR or Northern blot analysis, using appropriate probes for the transcripts. Reverse transcription-polymerase chain reaction (RT-PCR) can be used to analyze RNA samples containing mixtures of spliced and unspliced mRNA transcripts. For example, fluorescent-tagged primers annealing to polynucleotide sequences flanking the intron region are used to amplify both spliced and

unspliced target sequences of the transcript. The resultant amplification products are then separated by gel electrophoresis and quantitated by measuring the fluorescent emission of the appropriate band(s). Alternatively, a quantitative PCR approach can be used to quantitate the different amplification products. A comparison is made to determine the amount of spliced and unspliced transcripts present in the RNA sample. Suitable splice donor and splice acceptor sites leading to correctly and efficiently spliced transcripts and thus to polypeptides of the expected sequence only at high level may also be determined by more indirect means by assessing product purity and heterogeneity, e.g. by Western Blot analysis or mass spectroscopy methods (e.g. MS, ESI, MALDI, LC/MS).

A "heterologous intron" according to the invention is an intron placed at a sequence position within an exon coding for at least part of a signal peptide, meaning different from the intron position(s) in the native eukaryotic gene, or into a eukaryotic, prokaryotic or synthetic gene which naturally does not contain an intron within the signal peptide sequence. For example, the intron is inserted within an immunoglobulin exon encoding part of a signal peptide sequence, whereby the 5' and 3' ends of said exon are defined as occurring in a corresponding native immunoglobulin gene. In a native immunoglobulin gene of for example mouse or human origin the signal peptide sequence is coded by 2 exons whereby the 3' end of the exon encoding the N-terminal part of the signal peptide sequence is the first nucleotide of the codon encoding the amino acid at position -4 (counting backwards from the 3' end of the amino acid sequence of the signal peptide) and the 5' end of the second exon encoding the C-terminal part of the signal peptide sequence is the second nucleotide of the aforementioned codon. Not limiting examples for native genes without an intron in the signal peptide sequence are the intercellular adhesion molecule, monocyte chemoattractant protein 1, acid alpha-glucosidase, tumor necrosis factor receptor 2 and placental alkaline phosphatase, all of human origin. An intron useful in the constructs of this invention will generally be an intron, which improves the expression of the gene of interest, especially genes encoding antibodies and antibody derived fragments, when placed within the nucleotide sequence encoding the signal peptide sequence compared to cells transfected with the same vectors but encoding the cDNA of this gene of interest with the coding sequence having the identical polynucleotide sequence (natural, modified, optimized, partially optimized or non-optimized) but without any intron within the transcription unit. Furthermore, an intron useful in this invention will be an intron which leads to an at least comparable or improved expression of the gene of interest, especially genes encoding antibodies and antibody derived fragments, when placed in the nucleotide sequence encoding the signal peptide sequence of this gene compared to cells transfected with the same vectors but encoding the genomic DNA of this gene of interest with the coding sequence having the identical polynucleotide sequence (natural, modified, optimized, partially optimized or non-optimized) and in which the natural, modified, chimeric or synthetic intron(s) in the genomic DNA are in their natural sequence positions. The genomic DNA(s) to which the construct of the invention is compared to might thereby contain introns in one, two, more or all of their native positions normally also found in the natural or part of the natural gene(s). The heterologous intron can be placed into the signal peptide sequence of a native eukaryotic gene which might still contain introns in one, two, more or all of their native positions normally also found in the natural or part of the natural gene or into the cDNA of a eukaryotic, prokaryotic or

synthetic gene which does not contain an intron, whereby either identical heterologous introns or different heterologous introns can be used. For expression of heteromeric proteins either identical or different introns can be placed into the signal peptide sequences of the genes encoding the different polypeptides of a protein. Alternatively, an intron can be placed in just one or a few signal peptide sequences of the genes or all but one of the genes encoding the different polypeptides of a protein. Most preferred is the use of a single heterologous intron within a signal peptide sequence of a gene and no further introns to reduce the risk of alternative splicing and to reduce the size of the polynucleotide sequence.

The heterologous intron can be a) a natural intron sequence derived from the gene of interest itself, b) derived from a natural intron from the gene of interest itself but modified by nucleotide substitutions, deletions and/or insertions, c) a natural intron from a different gene, d) derived from the natural intron from a different gene but modified by nucleotide substitutions, deletions and/or insertions, e) a chimeric intron composed of different intron sequences derived from one or more natural intron sequences of the gene of interest and/or of different genes, f) a de novo designed synthetic intron or g) any combination of the above.

Again, an intron becomes a heterologous intron by virtue of its introduction at a sequence position within an exon, which naturally does not contain an intron at that sequence position. Sometimes this is also referred to as "non-natural placement" of the intron, thereby becoming a heterologous intron.

An intron useful in the constructs of this invention will generally be an efficient intron characterized by a splicing efficiency which results in all of the transcripts diverted to expression of the desired product. The efficient intron preferably has a splicing efficiency of about 90%-99%, preferably about 95-99%. Intron splicing efficiency is readily determined by quantifying the spliced transcripts versus the full-length, unspliced transcripts that contain the intron(s), using methods known in the art as described above.

Preferably, the heterologous intron present in the constructs of the invention has efficient splice donor and acceptor sites, as defined above, such that splicing of the primary transcript occurs at a frequency greater than 90%, preferably at least 95% and even more preferred at least 99%. In this manner, at least 99% of the transcripts will be translated into desired product.

Preferably, a heterologous intron present in the constructs of the invention has stop codons in all possible 3 reading frames and/or has a nucleotide sequence length which is not dividable by 3 to prevent a complete read through of the heterologous intron sequence in case of a non-splicing event. Furthermore, the intron contains a conserved branch site in the 3' region of the intron to allow for more efficient splicing. Ideally the intron sequence contains also a single restriction site close to the 5' end and/or the 3' end of the intron for subcloning purposes.

One intron suitable for use in the present invention is the intron from the human kappa light chain gene. In one embodiment, the intron from the human kappa light chain gene is further modified to introduce a) a single BglII restriction site close to the 5' end of the intron for cloning purposes, b) sequences which can act as stop codons in case of a non-splicing event of the messenger RNA which would then lead to premature translation termination of the protein and c) a conserved branch site in the 3' region of the intron to allow for more efficient splicing (SEQ ID NO: 1). In another embodiment, the intron used is the native intron from the hamster

## 11

dihydrofolate reductase gene (SEQ ID NO:2). In a further embodiment, the intron used is a intron from the hamster dihydrofolate reductase gene further modified to introduce a single BglII restriction site close to the 5' end of the intron for cloning purposes and optimized splice donor and acceptor sequences (SEQ ID NO:3). The intron is inserted within the genes of interest using any of the various known methods for modifying a nucleic acid in vitro or as described in example 1. More than one intron, either identical or different ones, can be inserted within a gene of interest. If convenient restriction sites are lacking within the genes of interest, they can be introduced using linkers and oligonucleotides by PCR, ligation or restriction and annealing. Alternatively, the heterologous intron-containing gene sequences can be prepared synthetically using various methods in organic chemistry.

A preferred intron is the nucleotide sequence comprising the sequence of SEQ ID NO:1, SEQ ID NO: 2 or SEQ ID NO:3 or modifications thereof which improve the expression of the gene of interest when placed in the signal peptide sequence of a transcription unit compared to cells transfected with the same vectors but encoding the cDNA of this gene of interest with the coding sequence having the identical polynucleotide sequence (native, modified, optimized, partially optimized or non-optimized) but without any intron within the transcription unit. Or the above introns lead as single intron to an at least comparable or even more preferred an improved expression of the gene of interest when placed in the signal peptide sequence of a transcription unit of this gene compared to cells transfected with the same vectors but encoding the genomic DNA of this gene of interest with the coding sequence having the identical polynucleotide sequence (native, modified, optimized, partially optimized or non-optimized) and in which the native, modified, chimeric or synthetic intron(s) in the genomic DNA are in their natural sequence positions. The genomic DNAs to which the construct of the invention is compared to might thereby contain introns in one, two, more or all positions normally also found in the native gene. Or the above introns lead to an improved expression of the gene of interest when placed in a signal peptide encoding exon of a transcription unit of this gene in addition to the introns in one, two, more or all positions normally also found in the native gene.

In a "comparative assay" the (heterologous) intron to be tested is introduced in the signal peptide sequence of a transcription unit of the gene(s) of interest. The transcription units are cloned into expression vectors. These recombinant expression vectors are subsequently introduced into the test cells, e.g. CHO-DG44, by transfection and the influence of the (heterologous) intron in question on the expression level of the gene of interest is determined for example by measuring the protein content of the gene of interest. Expression is compared to cells transfected with the same vectors but encoding either the cDNA of the corresponding gene(s) of interest with the coding sequence having the identical polynucleotide sequence (natural, modified, optimized, partially optimized or non-optimized) but without any intron within the transcription unit or the corresponding genomic DNA of this gene(s) of interest with the coding sequence having the identical polynucleotide sequence (native, modified, optimized, partially optimized or non-optimized) and in which native, modified, chimeric or synthetic intron(s) in the genomic DNA are in one, two, more or all of their native sequence positions. A corresponding test is described in examples 2, 3, 4 and 5 of the present invention.

For positioning of an intron within the native signal peptide encoding exons of a transcription unit the nucleotide sequence successions shown in Table 1 encoding the indi-

## 12

cated amino acids are preferred. If in the native polynucleotide sequence the amino acid pairs are not encoded by the preferred nucleotide successions the necessary adjustments in the nucleotide sequence can be achieved for example by performing site-specific mutagenesis, polymerase chain reaction mediated mutagenesis or de novo synthesis. Preferred are the nucleotide successions CAG:C, CAG:T, AAG:C, AAG:T, TAG:T or TAG:C whereby the colon denotes the site of intron insertion and even more preferred CAG:C, CAG:T, AAG:C or AAG:T. However, other sequences with sufficient splicing efficiency can be used as well. Intron sequences can be placed within an amino acid encoding codon or between two amino acid encoding codons. The heterologous introns can also be placed within the 5' or 3' untranslated region of an exon within a transcription unit, most preferably into the 5' untranslated region. In this case the heterologous intron is preferably inserted within the nucleotide sequences CAG:C, CAG:T, AAG:C, AAG:T, TAG:T or TAG:C whereby the colon denotes the site of intron insertion and even more preferred in CAG:C, CAG:T, AAG:C or AAG:T.

TABLE 1

| Preferred sites for positioning of heterologous intron |        |        |        |        |        |
|--|--------|--------|--------|--------|--------|
| CAG:C  | CAG:T  | AAG:C  | AAG:T  | TAG:C  | TAG:T  |
| GlnLeu   | GlnPhe | LysLeu | LysPhe | PheSer | PheSer |
| CAGCTN   | CAGTTY | AAGCTN | AAGTTY | TTTAGC | TTTAGT |
| GlnPro   | GlnLeu | LysPro | LysLeu | SerSer | SerSer |
| CAGCCN   | CAGTTR | AAGCCN | AAGTTR | TCTAGC | TCTAGT |
| GlnHis   | GlnSer | LysHis | LysSer | TyrSer | TyrSer |
| CAGCAY   | CAGTCN | AAGCAY | AAGTCN | TATAGC | TATAGT |
| GlnGln   | GlnTyr | LysGln | LysTyr | CysSer | CysSer |
| CAGCAR   | CAGTAY | AAGCAR | AAGTAY | TGTAGC | TGTAGT |
| GlnArg   | GlnCys | LysArg | LysCys | LeuSer | LeuSer |
| CAGCGN   | CAGTGY | AAGCGN | AAGTGY | CTTAGC | CTTAGT |
| SerAla   | GlnTrp | GlnAla | LysTrp | ProSer | ProSer |
| TCAGCN   | CAGTGG | CAAGCN | AAGTGG | CCTAGC | CCTAGT |
| ProAla   | SerVal | LysAla | GlnVal | HisSer | HisSer |
| CCAGCN   | TCAGTN | AAAGCN | CAAGTN | CATAGC | CATAGT |
| ThrAla   | ProVal | GluAla | LysVal | ArgSer | ArgSer |
| ACAGCN   | CCAGTN | GAAGCN | AAAGTN | CGTAGC | CGTAGT |
| AlaAla   | ThrVal | LeuSer | GluVal | IleSer | IleSer |
| GCAGCN   | ACAGTN | YTAAGC | GAAGTN | ATTAGC | ATTAGT |
| PheSer   | AlaVal | SerSer | LeuSer | ThrSer | ThrSer |
| TTCAGC   | GCAGTN | TCAAGC | YTAAGT | ACTAGC | ACTAGT |
| SerSer   | PheSer | ProSer | SerSer | AsnSer | AsnSer |
| TCCAGC   | TTCAGT | CCAAGC | TCAAGT | AATAGC | AATAGT |
| TyrSer   | SerSer | GlnSer | ProSer | SerSer | SerSer |
| TACAGC   | TCCAGT | CAAAGC | CCAAGT | AGTAGC | AGTAGT |
| CysSer   | TyrSer | ArgSer | GlnSer | ValSer | ValSer |



13

TABLE 1-continued

| Preferred sites for positioning of hetero-<br>logous intron |        |        |        |        |        |
|---|--------|--------|--------|--------|--------|
| CAG:C   | CAG:T  | AAG:C  | AAG:T  | TAG:C  | TAG:T  |
| TGCAGC  | TACAGT | MGAAGC | CAAAGT | GTTAGC | GTTAGT |
| LeuSer  | CysSer | IleSer | ArgSer | AlaSer | AlaSer |
| CTCAGC  | TGCAGT | ATAAGC | MGAAGT | GCTAGC | GCTAGT |
| ProSer  | LeuSer | ThrSer | IleSer | AspSer | AspSer |
| CCCAGC  | CTCAGT | ACAAGC | ATAAGT | GATAGC | GATAGT |
| HisSer  | ProSer | LysSer | ThrSer | GlySer | GlySer |
| CACAGC  | CCCAGT | AAAAGC | ACAAGT | GGTAGC | GGTAGT |
| ArgSer  | HisSer | ValSer | LysSer | LeuAla | LeuVal |
| CGCAGC  | CACAGT | GTAAGC | AAAAGT | YTAGCN | YTAGTN |
| IleSer  | ArgSer | AlaSer | ValSer | IleAla | IleVal |
| ATCAGC  | CGCAGT | GCAAGC | GTAAGT | ATAGCN | ATAGTN |
| ThrSer  | IleSer | GluSer | AlaSer | ValAla | ValVal |
| ACCAGC  | ATCAGT | GAAAGC | GCAAGT | GTAGCN | GTAGTN |
| AsnSer  | ThrSer | GlySer | GluSer |        |        |
| AACAGC  | ACCAGT | GGAAGC | GAAAGT |        |        |
| SerSer  | AsnSer |        | GlySer |        |        |
| AGCAGC  | AACAGT |        | GGAAGT |        |        |
| ValSer  | SerSer |        |        |        |        |
| GTCAGC  | AGCAGT |        |        |        |        |
| AlaSer  | ValSer |        |        |        |        |
| GCCAGC  | GTCAGT |        |        |        |        |
| AspSer  | AlaSer |        |        |        |        |
| GACAGC  | GCCAGT |        |        |        |        |
| GlySer  | AspSer |        |        |        |        |
| GGCAGC  | GACAGT |        |        |        |        |
|   | GlySer |        |        |        |        |
|   | GGCAGT |        |        |        |        |

colon: site of intron insertion, N = any base, Y = C or T, R = A or G, M = C or A

The term "synthetic" used in connection with a polynucleotide sequence is a non-natural polynucleotide (or portion of a polynucleotide) that differs from a wildtype polynucleotide sequence. For example, a synthetic gene or intron (or portion of a gene or intron) may contain one or more nucleic acid sequences not contiguous in nature (chimeric sequences), and/or may encompass substitutions, insertions, and deletions and combinations thereof.

As used herein, the term "hybridizes under stringent conditions" describes conditions for hybridization and washing which are known to those skilled in the art. Generally, stringent conditions are selected to be about 5-10° C. lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. The  $T_m$  is the temperature (under defined ionic strength, pH and nucleic acid concentra-

14

tion) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium. Stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30° C. for short probes (e.g. 10 to about 50 nucleotides) and at least about 60° C. for long probes (e.g. greater than about 50 nucleotides). Exemplary stringent conditions include hybridization at 60 to 65° C. in a hybridization buffer with 5×SSC and washing at 42° C. with 0.2×SSC/0.1% SDS. A positive hybridization signal is at least 2 times above background hybridization. The terms "homology", "homologous", "identity", "identical", "sequence identity" or "homologous sequence" are used interchangeably. Methods for calculating "homology" or "identity" are well known in the art. For sequence comparison typically one sequence acts as a reference sequence to which test sequences are compared. The sequences are aligned for maximal correspondence. Gaps can be introduced in either of the nucleic acid sequences in the comparison for optimal alignment. Percent identity between two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps and the length of each gap which need to be introduced for optimal alignment of the two sequences. The comparison of sequences and determination of percent identity between two sequences can be accomplished using mathematical algorithms. Default program parameters can be used or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent identity for the test sequence(s) relative to the reference sequence, based on the designated or default program parameters. One example of an algorithm that is suitable for determining identity is the BLAST algorithm (Altschul et al., J. Mol. Biol. 215, 403-410, 1990; Gish et al., Nature Genetics 3, 266-272, 1993; Madden et al., Meth. Enzymol. 266, 131-141, 1996; Zhang et al., Genome Res. 7, 649-656, 1997; Altschul et al., Nucleic Acids Res. 25, 3389-3402, 1997). Other computerized implementations of alignment algorithms are GAP, PILEUP, BESTFIT, FASTA and TFASTA in the Wisconsin Genetics Software Package. However, percent identity can be also determined by manual alignment and visual inspection and calculation.

"Exons" are defined as expressed regions of an eukaryotic gene which remain after the removal of the (heterologous) introns via splicing from the messenger RNA (mRNA) precursor in the mature mRNA. Most exons contain only coding sequences but some exons are non-coding or partially coding exons. Usually they are positioned at the 5' or 3' end of an mRNA and contain untranslated, non-coding regions which might for example contribute enhancer or stabilization domains to the final transcript, resulting in increased translation of protein.

The term "vector" or "expression vector" as used herein relates to naturally occurring or synthetically generated constructs for uptake, proliferation, expression or transmission of nucleic acids in a cell, e.g. plasmids, minicircles, phagemids, cosmids, artificial chromosomes/mini-chromosomes, bacteriophages, viruses such as baculovirus, retrovirus, adenovirus, adeno-associated virus, herpes simplex virus, bacteriophages. Vectors can either integrate into the genome of the host cell or remain as autonomously replicating construct within the host cell. Methods used to construct vectors are well known to a person skilled in the art and described in various publications. In particular techniques for constructing suitable vectors, including a description of the functional and regulatory components such as promoters, enhancers, termination and polyadenylation signals, selection markers,

origins of replication, and splicing signals, are known to the person skilled in the art. The eukaryotic expression vectors will typically contain also prokaryotic sequences that facilitate the propagation of the vector in bacteria such as an origin of replication and antibiotic resistance genes for selection in bacteria which might be removed before transfection of eukaryotic cells. A variety of eukaryotic expression vectors, containing a cloning site into which a polynucleotide can be operably linked, are well known in the art and some are commercially available from companies such as Agilent Technologies, Santa Clara, Calif.; Invitrogen, Carlsbad, Calif.; Promega, Madison, Wis. or Invivogen, San Diego, Calif.

A preferred embodiment of the invention are vectors or polynucleotide sequences containing one or more transcription units encoding genes of interest which comprise at least one heterologous intron within a signal peptide sequence. Also preferred according to the invention are vectors or polynucleotide sequences comprising a signal peptide sequence, preferably an immunoglobulin signal peptide sequence, with a heterologous intron for improved expression and having suitable restriction sites within or after the 3' end of the signal peptide sequence to allow the cloning of the gene of interest via recognitions sequences for restriction endonuclease in frame with the signal peptide sequence thus building a functional transcription unit encoding the complete gene of interest. The restriction enzyme sites might also be placed within the heterologous intron sequence. Furthermore, the signal peptide sequence might already be flanked at its 3' end by sequences coding for the first amino acids of the variable region of an antibody or antibody-derived heavy or light chain or other polypeptide sequences.

The term "promoter" denotes a polynucleotide sequence which allows and controls the transcription of the genes or sequences operably connected therewith. A promoter contains recognition sequences for binding RNA polymerase and the initiation site for transcription (transcription initiation site). In order to express a desired sequence in a certain cell type or a host cell a suitable functional promoter must be chosen. A large number of promoters, including constitutive, inducible and repressible promoters from a variety of different sources, are well known in the art (and identified in databases such as GenBank) and are available as separate elements or elements cloned within polynucleotide sequences from commercial (e.g. depositories such as ATCC as well as other commercial sources) or individual sources. In inducible promoters the activity of the promoter may be increased or reduced in response to a signal. For example, the tetracycline (tet) promoter containing the tetracycline operator sequence (tetO) can be induced by a tetracycline-regulated transactivator protein (tTA). Binding of the tTA to the tetO is inhibited in the presence of tet. Examples for other inducible promoters are jun, fos, metallothionein and heat shock promoters. Of the promoters which are particularly suitable for high expression in eukaryotes, there are for example the ubiquitin/S27a promoter of the hamster (WO 97/15664), SV 40 early promoter, adenovirus major late promoter, mouse metallothionein-I promoter, the long terminal repeat region of Rous Sarcoma Virus, the early promoter of human Cytomegalovirus (CMV). Examples of other heterologous mammalian promoters are the actin, immunoglobulin or heat shock promoter(s).

The aforementioned promoters are well known in the art. A corresponding heterologous promoter can be functionally connected to other regulatory sequences in order to increase/regulate the transcription activity in an expression cassette. For example, the promoter may be functionally linked to enhancer sequences in order to increase the transcriptional

activity. For this, one or more enhancers and/or several copies of an enhancer sequence may be used, e.g. a CMV or SV40 enhancer. Accordingly, an expression vector according to the invention, in another embodiment, contains one or more enhancers/enhancer sequences, preferably a CMV or SV40 enhancer.

The term "enhancer" denotes a polynucleotide sequence which in the cis location acts on the activity of a promoter and thus stimulates the transcription of a gene or coding sequence functionally connected to this promoter. Unlike promoters the effect of enhancers is independent of position and orientation and they can therefore be positioned in front of or behind a transcription unit, within an intron or even within the coding region. The enhancer may be located both in the immediate vicinity of the transcription unit and at a considerable distance from the promoter. It is also possible to have a physical and functional overlap with the promoter. The skilled artisan will be aware of a number of enhancers from various sources (and deposited in databanks such as GenBank, e.g. SV40 enhancers, CMV enhancers, polyoma enhancers, adenovirus enhancers) which are available as independent elements or elements cloned within polynucleotide sequences (e.g. deposited at the ATCC or from commercial and individual sources). A number of promoter sequences also contain enhancer sequences such as the frequently used CMV promoter. The human CMV enhancer is one of the strongest enhancers identified hitherto. One example of an inducible enhancer is the metallothionein enhancer, which can be stimulated by glucocorticoids or heavy metals.

"Transcription-regulatory elements" normally comprise a promoter upstream of the gene sequence to be expressed, transcription initiation and termination sites and a polyadenylation signal.

The term "transcription initiation site" refers to a nucleic acid in the construct corresponding to the first nucleic acid incorporated into the primary transcript, i.e. the mRNA precursor. The transcription initiation site may overlap with the promoter sequences.

The term "transcription termination site" or "transcription termination element" refers to a nucleotide sequence normally represented at the 3' end of the gene of interest or of the stretch of sequences to be transcribed, that causes RNA polymerase to terminate transcription.

The "polyadenylation signal" or "polyA" is a signal sequence which causes cleavage at a specific site at the 3' end of the eukaryotic mRNA and post-transcriptional incorporation of a sequence of about 100-200 adenine nucleotides (SEQ ID NO: 28) (polyA tail) at the cleaved 3' end. The polyadenylation signal comprises the sequence AATAAA about 10-30 nucleotides upstream of the cleavage site and a sequence located downstream. Various polyadenylation elements are known such as tk polyA, SV40 late and early polyA, BGH polyA (described for example in U.S. Pat. No. 5,122,458) or hamster growth hormone polyA (WO2010010107).

A "transcription unit", "expression unit" or "expression cassette" defines a region within a vector, construct or polynucleotide sequence that contains one or more genes to be transcribed, wherein the genes contained within the segment are operably linked to each other. They are transcribed from a single promoter and transcription is terminated by at least one polyadenylation signal. As a result, the different genes are at least transcriptionally linked. More than one protein or product can be transcribed and expressed from each transcription unit (multicistronic transcription unit). Each transcription unit will comprise the regulatory elements necessary for the transcription and translation of any of the selected sequence

that are contained within the unit. And each transcription unit may contain the same or different regulatory elements. For example, each transcription unit may contain the same terminator. IRES element or introns may be used for the functional linking of the genes within a transcription unit. A vector or polynucleotide sequence may contain more than one transcription unit.

"Translation regulatory elements" comprise a translation initiation site (AUG), a stop codon and a polyA signal for each individual polypeptide to be expressed. An internal ribosome entry site (IRES) may be included in some constructs. In order to optimize expression it may be advisable to remove, add or alter 5'- and/or 3'-untranslated regions of the nucleic acid sequence to be expressed to eliminate any potentially extra inappropriate alternative translation initiation codons or other sequences that may interfere with or reduce expression, either at the level of transcription or translation. Consensus ribosome binding sites (Kozak sequence) can be inserted immediately upstream of the start codon to enhance translation and thus expression. Increased A/U contents around this ribosome binding site further a more efficient ribosome binding.

To produce a secreted polypeptide the gene of interest usually includes a "signal sequence" or "signal peptide sequence" encoding a "leader" or "signal peptide" that directs the newly synthesized polypeptide to and through the ER membrane where the polypeptide can be routed for secretion. The leader or signal peptide is often but not universally at the amino terminus of a secreted protein and is cleaved off by signal peptidases after the protein crosses the ER membrane. Signal peptides may also be called targeting signals, signal sequences, transit peptides, or localization signals. Signal peptides show great variance in overall length (approximately 15-50 amino acids) and amino acid sequence. All contain a hydrophobic region preceded by a domain with basic amino acids and followed by a slightly polar C-terminal domain. The latter often contains helix breaking proline or glycine residues as well as uncharged residues in positions -3 and -1 that determine the site of cleavage. The gene sequence will generally, but not necessarily, contain its own signal peptide sequence. Where the native signal peptide sequence is absent, a heterologous signal peptide sequence can be fused to the selected sequence. Or the native signal peptide sequence can be replaced by a heterologous one. Numerous signal peptide sequences are known to the skilled artisan and deposited in sequence databanks such as GenBank and EMBL. Signal peptides can direct proteins also to other organelles such as the nucleus, mitochondrial matrix, chloroplast, apoplast and peroxisome.

An "internal ribosome entry site" or "IRES" describes a sequence which functionally promotes translation initiation independent from the gene 5' of the IRES and allows two cistrons (open reading frames) to be translated from a single transcript in an animal cell. The IRES provides an independent ribosome entry site for translation of the open reading frame immediately downstream of it. Unlike bacterial mRNA which can be polycistronic, i.e., encode several different polypeptides that are translated sequentially from the mRNAs, most mRNAs of animal cells are monocistronic and code for the synthesis of only one polypeptide. With a polycistronic transcript in a eukaryotic cell, translation would initiate from the 5' most translation initiation site, terminate at the first stop codon, and the transcript would be released from the ribosome, resulting in the translation of only the first encoded polypeptide in the mRNA. In a eukaryotic cell, a polycistronic transcript having an IRES operably linked to the second or subsequent open reading frame in the transcript

allows the sequential translation of that downstream open reading frame to produce the two or more polypeptides encoded by the same transcript. The IRES can be of varying length and from various sources, e.g. encephalomyocarditis virus (EMCV), picornavirus (e.g. FMDV), polio virus (PV), or hepatitis C virus (HCV). Various IRES sequences and their use in vector construction have been described and are well known in the art. The downstream coding sequence is operably linked to the 3' end of the IRES at any distance that will not negatively affect the expression of the downstream gene. The optimum or permissible distance between the IRES and the start of the downstream gene can be readily determined by varying the distance and measuring expression as a function of the distance.

The terms "gene", "gene of interest", "desired sequence", "polynucleotide of interest" or "desired gene" as used herein have the same meaning and refer to a polynucleotide sequence of any length that encodes a product of interest. The gene may further comprise regulatory sequences preceding (5' non-coding or untranslated sequences) and following (3' non-coding or untranslated sequences) the coding sequence. The selected sequence can be full length or a truncated gene, a fusion or tagged gene, and can be a cDNA, a genomic DNA, or a DNA fragment. It is generally understood that genomic DNA encoding for a polypeptide or RNA includes non-coding regions (i.e. introns) that are spliced from mature messenger RNA (mRNA) and are therefore not present in cDNA encoding for the same polypeptide or RNA. It can be the native sequence, i.e. naturally occurring form(s), or can be mutated, or comprising sequences derived from different sources or otherwise modified as desired. These modifications include codon optimizations to optimize codon usage in the selected host cell or tagging. Furthermore they can include removal or additions of cis-acting sites such as (cryptic) splice donor, acceptor sites and branch points, polyadenylation signals, TATA-boxes, chi-sites, ribosomal entry sites, repeat sequences, secondary structures (e.g. stem loops), binding sites for transcription factors or other regulatory factors, restriction enzyme sites etc. to give just a few, but not limiting examples. The selected sequence can encode a secreted, cytoplasmic, nuclear, membrane bound or cell surface polypeptide.

"Open reading frame" or "ORF" refers to a length of nucleic acid sequence, either DNA, cDNA or RNA, that comprises a translation start signal or initiation codon, such as an ATG or AUG, and a termination codon and can be potentially translated into a polypeptide sequence.

Within the scope of the present description the terms "functional linking", "functionally linked" or "operably linked" means that two or more nucleic acid sequences or sequence elements are positioned in a way that permits them to function in their intended manner. For example, a promoter/enhancer or terminator is functionally linked to a coding gene sequence if it is able to control or modulate the transcription of the linked gene sequence in the cis position. Generally, but not necessarily, the DNA sequences that are functionally linked are contiguous and, where necessary to join two polypeptide coding regions or in the case of a secretion signal peptide, contiguous and in reading frame. However, although an operably linked promoter is generally located upstream or an operably linked terminator is generally located downstream of the coding sequence, it is not necessarily contiguous with it. Enhancers do not have to be contiguous as long as they increase the transcription of the coding sequence. For this they can be located upstream or downstream of the coding sequence and even at some distance. A polyadenylation site is operably linked to a coding sequence if it is located at the 3'

end of the coding sequence in a way that transcription proceeds through the coding sequence into the polyadenylation signal. Linking is accomplished by recombinant methods known in the art, e.g. using PCR methodology, by ligation at suitable restriction sites or by annealing. Synthetic oligonucleotide linkers or adaptors can be used in accord with conventional practice if suitable restriction sites are not present.

The term “nucleic acid”, “nucleic acid sequence”, “nucleotide sequence”, “polynucleotide”, “polynucleotide sequence”, “RNA sequence” or “DNA sequence” as used herein refers to an oligonucleotide, nucleotide or polynucleotide and fragments and portions thereof and to DNA or RNA of genomic or synthetic origin, which may be single or double stranded and represent the sense or antisense strand. The sequence may be a non-coding sequence, a coding sequence or a mixture of both. The nucleic acid sequences of the present invention can be prepared using standard techniques well known to one of skill in the art.

The term “encoding” or “coding” refers to the inherent property of specific sequences of nucleotides in a nucleic acid, such as a gene in chromosome or an mRNA, to serve as templates for in vitro or in vivo synthesis of other polymers and macromolecules in biological processes having a defined sequence of nucleotides (i.e. rRNA, tRNA, other RNA molecules) or amino acids and the biological properties resulting therefrom. Accordingly, a gene codes for a protein if the desired protein is produced in a cell or another biological system by transcription and subsequent translation of the mRNA whereby the boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxyl) terminus. Both the coding strand, the nucleotide sequence of which is identical to the mRNA sequence and is usually provided in sequence listings of databanks, e.g. EMBL or GenBank, and non-coding strand, used as the template for the transcription, of a gene or cDNA can be referred to as encoding the protein or other product of that gene or cDNA. A nucleic acid that encodes a protein includes any nucleic acids that have different nucleotide sequences but encode the same amino acid sequence of the protein due to the degeneracy of the genetic code. Nucleic acids and nucleotide sequences that encode proteins may include introns. In the Sequence Listing the sequences are presented as DNA rather than RNA sequence. For example, when presented as DNA the start codon is presented as ATG rather than AUG.

The term “cDNA” in the context of this invention refers to deoxyribonucleic acids produced by reverse transcription and typically second-strand synthesis of mRNA or other RNA produced by a gene. It can also be generated by de novo synthesis. If double-stranded, a cDNA molecule has both a coding or sense and a non-coding or antisense strand.

“Chemically synthesized” or “de novo synthesized” as related to a DNA sequence, means that the component nucleotides were assembled in vitro. Manual chemical synthesis of DNA may be accomplished using well-established procedures, or automated chemical synthesis can be performed using one of a number of commercially available machines. Furthermore, if needed the genes can be tailored for optimal gene expression based on optimization of nucleotide sequence to reflect the codon bias of the host cell. The skilled artisan appreciates the likelihood of successful gene expression if codon usage is biased towards those codons favored by the host. Determination of preferred codons can be based on a survey of genes derived from the host cell where sequence information is available. Furthermore, potential cryptic splice sites, direct repeats, secondary structure elements and other

motifs interfering with expression can be removed and the GC content optimized to improve RNA stability.

The terms “restriction endonuclease” and “restriction enzyme” are used interchangeably and refer to an enzyme that binds and cuts within a specific nucleotide sequence within double stranded DNA.

“Polymerase chain reaction” or “PCR” is an in vitro method for enzymatically amplifying specific nucleic acid sequences. Generally, knowledge of the sequence from the ends of the region of interest or beyond is needed such that oligonucleotide primers can be designed. These primers will be fully or in part identical or similar in sequence to opposite strands on the template to be amplified. PCR involves a repetitive series of temperature cycles with each cycle comprising three stages: denaturation of the template nucleic acid to separate the strands of the target molecule, annealing of single stranded oligonucleotide primer(s) to the template nucleic acid, and extension of the annealed primer(s) by DNA polymerase.

The term “expression” as used herein refers to transcription and/or translation of a heterologous nucleic acid sequence within a host cell. The level of expression of a desired product in a host cell may be determined on the basis of either the amount of corresponding RNA or mRNA that is present in the cell, or the amount of the desired polypeptide encoded by the selected sequence. For example, mRNA transcribed from a selected sequence can be quantitated by Northern blot hybridization, ribonuclease RNA protection, in situ hybridization to cellular RNA or by PCR. Proteins encoded by a selected sequence can be quantitated by various methods, e.g. by ELISA, by Western blotting, by radioimmunoassays, by immunoprecipitation, by assaying for the biological activity of the protein, or by immunostaining of the protein followed by FACS analysis PCR.

The term “polypeptide” is used interchangeably with “amino acid residue sequence”, “amino acid sequence” or the term “protein” and refers to polymers of amino acids of any length. These terms also include proteins that are post-translationally modified through reactions that include, but are not limited to glycosylation, glycation, acetylation, phosphorylation, oxidation, amidation or protein processing. Modifications and changes, for example fusions to other proteins, amino acid sequence substitutions, deletions or insertions, can be made in the structure of a polypeptide while the molecule maintains its biophysical properties and/or biological functional activity. For example certain amino acid sequence substitutions can be made in a polypeptide or its underlying nucleic acid coding sequence and a protein can be obtained with like properties. Furthermore, modifications and changes can be made in the structure of a polypeptide to gain or improve biophysical properties and/or to modulate, introduce or abrogate biological functional activity. For example, modification and changes can be introduced to improve the solubility and/or stability of a protein, such as introducing the amino substitution Ser228Pro in the hinge region of an IgG4 molecule to stabilize the intermolecular disulfide bridge between the heavy chains and thus reducing the occurrence of half molecules. Or in case of IgGs amino acid substitutions can be made to modulate binding to one or more Fc gamma receptors to either improve or abrogate antibody dependent cellular cytotoxicity, or to improve binding to the neonatal Fc receptor for prolonged half-life or to improve or abrogate binding to components involved in the complement dependent cytotoxicity. Amino acid modifications can be prepared for example by performing site-specific mutagenesis or polymerase chain reaction mediated mutagenesis on its underlying nucleic acid sequence. The term “polypeptide” thus also

includes, for example, fusion proteins consisting of an immunoglobulin component, e.g. the Fc component, and a growth factor, e.g. an interleukin. In addition, the polypeptides may multimerise and form homo- or heteromers.

"Immunoglobulins", or "antibodies" are proteins selected from among the globulins, which are formed as a reaction of the host organism to a foreign substance (=antigen) from differentiated B-lymphocytes (plasma cells). They serve to defend specifically against these foreign substances. There are various classes of immunoglobulins: IgA, IgD, IgE, IgG, IgM, IgY, IgW. The terms immunoglobulin and antibody are used interchangeably. As used herein, the term "immunoglobulin" or "antibody" includes a polyclonal, monoclonal, monospecific, bi-specific, multi-specific, a single chain antibody, an antigen-binding fragment of an antibody (e.g., an Fab or F(ab')<sub>2</sub> fragment), a disulfide-linked Fv, etc. Antibodies can be of any species and include chimeric and humanized antibodies. "Chimeric" antibodies are molecules in which antibody domains or regions are derived from different species. For example the variable region of heavy and light chain can be derived from rat or mouse antibody and the constant regions from a human antibody. In "humanized" antibodies only minimal sequences are derived from a non-human species. Often only the CDR amino acid residues of a human antibody are replaced with the CDR amino acid residues of a non-human species such as mouse, rat, rabbit or llama. Sometimes a few key framework amino acid residues with impact on antigen binding specificity and affinity are also replaced by non-human amino acid residues. Antibodies may be produced through chemical synthesis, via recombinant or transgenic means, via cell (e.g., hybridoma) culture, or by other means.

Immunoglobulins are tetrameric polypeptides composed of two pairs of a heterodimer each formed by a heavy and light chain. Stabilization of both the heterodimers as well as the tetrameric polypeptide structure occurs via interchain disulfide bridges. Each chain is composed of structural domains called "immunoglobulin domains" or "immunoglobulin regions" whereby the terms "domain" or "region" are used interchangeably. Each domain contains about 70-110 amino acids and forms a compact three-dimensional structure. Both heavy and light chain contain at their N-terminal end a "variable domain" or "variable region" with less conserved sequences which is responsible for antigen recognition and binding. The variable region of the light chain is also referred to as "VL" and the variable region of the heavy chain as "VH". The variable regions exhibit the same general structure of 4 relatively conserved framework regions, also referred to as "FR1" to "FR4", interspersed by three hypervariable regions, also called complementarity determining regions ("CDR"), which are the main contributors to the specific binding of an antigen. CDR and FR residues are determined according to the standard sequence definition (Kabat et al. (1991), "Sequences of proteins of immunological interest", US Dept. Health and Human Services) and a structural definition (Chothia and Les, J. Mol. Biol. 196, 901-917, 1987). The substructure of a VH or VL region can be described as follows: FR1-CDR1-FR2-CDR2-FR3-CDR3-FR4. The C-terminal part of each chain contains one or more domains with much more conserved "constant" sequences, also called "constant domains" or "constant regions". An immunoglobulin light chain, kappa or lambda type, contains a single constant domain, also referred to as "CL" domain or region. The constant domains of an immunoglobulin heavy chain are called "CH" domains or regions. Depending on the immunoglobulin class the number of domains varies between 3 and 4. The heavy chain of an IgG class immunoglobulin with the

subclasses IgG1, IgG2, IgG3 and IgG4 contains 3 different CH domains: "CH1", "CH2" and "CH3". The immunoglobulin fragments composed of the CH2 and CH2 domains of the heavy chain are called "Fc fragments", "Fc region" or "Fc" because of their crystallization propensity (Fc=fragment crystallizable). Between CH1 and CH2 a region with more variation with regard to sequence and length is found, the "hinge region" also referred to as "hinge" or "H". It not only stabilizes the heterodimers by interchain disulfide bridges between the heavy chains but it also determines the flexibility between both Fab arms and the flexibility between the Fab arms and the Fc region of the immunoglobulin molecule. The substructure of an IgG heavy chain can be described as follows: CH1-H—CH2-CH3.

The term "antibody derived molecules" is used interchangeably with "antibody derived fragments" or "antibody fragments" and refers to polypeptides which contain only part(s) of one or more antibody domain(s) or region(s) and/or complete domain(s) or region(s). The antibody fragments can be either a) forming a molecule on their own, b) linked with each other in different combinations, c) fused to non-antibody sequences, d) fused or linked to non-polypeptide (e.g. radionucleotides) or d) any combination of the above. These polypeptides can exist either as monomers or as multimers whereby polypeptides can have identical or different sequences.

"Fab fragments" ("Fragment antigen-binding=Fab) or "Fab" consist of the variable regions of both antibody heavy and light chains (VH and VL) which are held together by the adjacent constant regions (CH1 and CL). These may be formed by protease digestion, e.g. with papain, from conventional antibodies, but similar Fab fragments may also be produced in the mean time by genetic engineering. Further antibody fragments include "F(ab')<sub>2</sub> fragments" or "F(ab')<sub>2</sub>", which may be prepared by proteolytic cleaving with pepsin or by genetic engineering in which both Fab arms of an antibody are still linked via inter-heavy chain disulfide bridges located within the hinge region.

The immunoglobulin fragments composed of the CH2 and CH3 domains of the antibody heavy chain are called "Fc fragments", "Fc region" or "Fc" because of their crystallization propensity (Fc=fragment crystallizable). These may be formed by protease digestion, e.g. with papain or pepsin from conventional antibodies but may also be produced by genetic engineering. The N-terminal part of the Fc fragment might vary depending on how many amino acids of the hinge region are still present.

The term "Fc-fusion protein" describes polypeptides which contain as a fusion partner a natural or modified (e.g. substitutions, deletions, insertions) Fc region of an immunoglobulin. Fc fusion proteins can be either naturally occurring proteins (e.g. antibodies) or engineered recombinant proteins (e.g. TNF receptor-Fc fusion protein or a VH region fused to an Fc region). The Fc-fusion proteins can exist either as monomers or as multimers whereby polypeptides can have identical or different sequences, might contain linker sequences between the two fusion partners and/or part of the hinge region or modified hinge regions or the polypeptide is fused directly to the CH2 domain.

Using genetic engineering methods it is possible to produce shortened antibody fragments which consist only of the variable regions of the heavy (VH) and of the light chain (VL). These are referred to as "Fv fragments" (Fragment variable=fragment of the variable part) or "Fv". Since these Fv-fragments lack the covalent bonding of the two chains by the cysteines of the constant chains, the Fv fragments are

often stabilised. It is advantageous to link the variable regions of the heavy and of the light chain by a short peptide fragment, e.g. of 10 to 30 amino acids, preferably 15 amino acids. In this way a single peptide strand is obtained consisting of VH and VL, linked by a peptide linker. An antibody protein of this kind is known as a "single-chain-Fv" or "scFv". Examples of scFv-antibody proteins of this kind are known from the prior art. In addition, more than one VH and/or VL region can be linked together. In addition, the polypeptides may multimerise and form homo- or heteromers.

In recent years, various strategies have been developed for preparing scFv as a multimeric derivative. This is intended to lead, in particular, to recombinant antibodies with improved pharmacokinetic and biodistribution properties as well as with increased binding avidity. In order to achieve multimerisation of the scFv, scFv were prepared as fusion proteins with multimerisation domains. The multimerisation domains may be, e.g. the CH3 region of an IgG or coiled coil structure (helix structures) such as Leucine-zipper domains. However, there are also strategies in which the interaction between the VH/VL regions of the scFv are used for the multimerisation (e.g. di-, tri- and pentabodies). By diabody the skilled person means a bivalent homodimeric scFv derivative. The shortening of the linker in an scFv molecule to 5-10 amino acids leads to the formation of homodimers in which an inter-chain VH/VL-superimposition takes place. Diabodies may additionally be stabilised by the incorporation of disulphide bridges. Examples of diabody-antibody proteins are known from the prior art.

By minibody the skilled person means a bivalent, homodimeric scFv derivative. It consists of a fusion protein which contains the CH3 region of an immunoglobulin, preferably IgG, most preferably IgG1 as the dimerisation region which is connected to the scFv via a Hinge region (e.g. also from IgG1) and a linker region. Examples of minibody-antibody proteins are known from the prior art.

By triabody the skilled person means a: trivalent homotrimeric scFv derivative. ScFv derivatives wherein VH-VL are fused directly without a linker sequence lead to the formation of trimers.

The skilled person will also be familiar with so-called miniantibodies which have a bi-, tri- or tetravalent structure and are derived from scFv. The multimerisation is carried out by di-, tri- or tetrameric coiled coil structures. In a preferred embodiment of the present invention, the gene of interest is encoded for any of those desired polypeptides mentioned above, preferably for a monoclonal antibody, a derivative or fragment thereof.

The "polypeptide of interest", "protein of interest" or "product of interest" includes proteins, polypeptides, fragments thereof, peptides, fusion proteins all of which can be expressed in the selected host cell. Desired proteins can be for example antibodies, enzymes, cytokines, lymphokines, adhesion molecules, receptors and derivatives or fragments thereof, and any other polypeptides that can serve as agonists or antagonists and/or have therapeutic or diagnostic use. Other proteins of interest are, for example, proteins/polypeptides, which are used to change the properties of host cells within the scope of so-called "Cell Engineering", such as e.g. anti-apoptotic proteins, chaperones, metabolic enzymes, glycosylation enzymes and the derivatives or fragments thereof, but are not restricted thereto.

Especially, desired proteins/polypeptides or proteins of interest are for example, but not limited to Fc receptors, enzymes, tumor necrosis factor receptor, growth hormone receptors, insulin, insulin-like growth factor, hGH, tPA, cytokines, such as interleukines (IL), e.g. IL-1, IL-2, IL-3,

IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IL-16, IL-17, IL-18, interferon (IFN) alpha, IFN beta, IFN gamma, IFN omega or IFN tau, tumor necrosis factor (TNF), such as TNF alpha and TNF beta, TNF gamma, TRAIL; G-CSF, GM-CSF, M-CSF, MCP-1, VEGF and nanobodies. Also included is the production of erythropoietin or any other hormone growth factors and any other polypeptides that can serve as agonists or antagonists and/or have therapeutic or diagnostic use. The method according to the invention can also be advantageously used for production of antibodies, such as monoclonal, polyclonal, multispecific and single chain antibodies, or fragments derived thereof, e.g. Fab, Fab', F(ab')2, Fc and Fc'-fragments, heavy and light immunoglobulin chains and their constant, variable or hyper-variable region as well as Fv- and Fd-fragments.

The "product of interest" may also be an antisense RNA, tRNA, rRNAs, other RNAs being part of riboproteins or other regulatory RNAs.

The method of the present invention may be performed in all eukaryotic cells. Cells and cell lines may be present e.g. in a cell culture and include but are not limited to eukaryotic cells, such as yeast, plant, insect or mammalian cells. For example, the cells may be oocytes, embryonic stem cells, hematopoietic stem cells or any type of differentiated cells. A method is preferred wherein the eukaryotic cell is a mammalian cell. More preferred is a method wherein the mammalian cell is a rodent cell. Furthermore, preferred is a method wherein the mammalian cell is a human, simian, murine, rat, rabbit, hamster, goat, bovine, sheep or pig cell. Preferred cell lines or "host cells" for the production of biopharmaceuticals are human, mice, rat, monkey, or rodent cell lines. More preferred are hamster cells, preferably BHK21, BHK TK<sup>-</sup>, CHO, CHO-K1, CHO-DUKX, CHO-DUKX B1, CHO-S and CHO-DG44 cells or the derivatives/progenies of any of such cell lines. Particularly preferred are CHO-DG44, CHO-DUKX, CHO-K1, CHO-S and BHK21, and even more preferred CHO-DG44 and CHO-DUKX cells. Furthermore, murine myeloma cells, preferably NS0 and Sp2/0 cells or the derivatives/progenies of any of such cell lines are also known as production cell lines for biopharmaceutical proteins.

Host cells are most preferred, when being established, adapted, and completely cultivated under serum free conditions, and optionally in media which are free of any protein/peptide of animal origin. Commercially available media such as Ham's F12 (Sigma, Deisenhofen, Germany), RPMI-1640 (Sigma), Dulbecco's Modified Eagle's Medium (DMEM; Sigma), Minimal Essential Medium (MEM; Sigma), Iscove's Modified Dulbecco's Medium (IMDM; Sigma), CD-CHO (Invitrogen, Carlsbad, Calif.), CHO-S—SFMII (Invitrogen), serum-free CHO Medium (Sigma), protein-free CHO Medium (Sigma), EX-CELL Media (SAFC), CDM4CHO and SFM4CHO (HyClone) are exemplary appropriate nutrient solutions. Any of the media may be supplemented as necessary with a variety of compounds examples of which are hormones and/or other growth factors (such as insulin, transferrin, epidermal growth factor, insulin like growth factor), salts (such as sodium chloride, calcium, magnesium, phosphate), buffers (such as HEPES), nucleosides (such as adenosine, thymidine), glutamine, glucose or other equivalent energy sources, antibiotics, trace elements. Any other necessary supplements may also be included at appropriate concentrations that would be known to those skilled in the art. In the present invention the use of serum-free medium is preferred, but media supplemented with a suitable amount of serum can also be used for the cultivation of host cells. For the

growth and selection of genetically modified cells expressing a selectable gene a suitable selection agent can be added to the culture medium.

The “transfection” of eukaryotic host cells with polynucleotide sequences or expression vectors, resulting in genetically modified cells, recombinant or transgenic cells, can be performed by any method well known to the skilled artisan. Transfection methods include but are not limited to liposome-mediated transfection, calcium phosphate co-precipitation, electroporation, polycation (e.g. DEAE dextran)-mediated transfection, protoplast fusion, microinjection and viral infections. Preferably, the transfection is a stable transfection. The transfection method that provides optimal transfection frequency and expression of the heterologous genes or polynucleotides in the particular host cell line and type is favored. Suitable methods can be determined by routine procedures. For stable transfectants the constructs are either integrated into the host cell's genome or an artificial chromosome/mini-chromosome or located episomally so as to be stably maintained within the host cell. For generation of genetically modified cells expressing the product(s) of interest all required heterologous genes can be located on a single vector or polynucleotide sequence in mono- or multicistronic transcription units. In this case the host cell is transfected with single vectors or polynucleotide sequences. The heterologous genes can also be positioned on different vectors or polynucleotide sequences. In this case host cells are either co-transfected with all vectors or polynucleotide sequences and/or are transfected in successive rounds with the vectors or polynucleotide sequences encoding the genes of interest. In case of co-transfection the ratios of the different vectors can be varied.

By definition, every polynucleotide sequence or every gene inserted in a host cell and the respective protein or RNA encoded thereby is referred to as “heterologous”, “heterologous sequence”, “heterologous gene”, “heterologous coding sequence”, “transgene” or “heterologous protein” with respect to the host cell. This applies even if the sequence to be introduced or the gene to be introduced is identical to an endogenous sequence or an endogenous gene of the host cell. For example, a hamster actin gene introduced into a hamster host cell is by definition a heterologous gene. The term “recombinant” is used exchangeably with the term “heterologous” throughout the specification of this present invention, especially in the context with protein expression. Thus, a “recombinant” protein is a protein expressed from a heterologous or recombinant polynucleotide sequence.

The term “selection marker gene” refers to a gene that only allows cells carrying the gene to be specifically selected for or against in the presence of a corresponding selection agent. By way of illustration, an antibiotic resistance gene can be used as a positive selectable marker gene that allows the host cell transformed with the gene to be positively selected for in the presence of the corresponding antibiotic; a non-transformed host cell would not be capable of growth or survival under the selection culture conditions. Selectable markers can be positive, negative or bifunctional. Positive selectable markers allow selection for cells carrying the marker by conferring resistance to a drug or compensate for a metabolic or catabolic defect in the host cell. In contrast, negative selection markers allow cells carrying the marker to be selectively eliminated. For example, using the HSV-tk gene as a marker will make the cells sensitive to agents such as acyclovir and gancyclovir. The selectable marker genes used herein, including the amplifiable selectable genes, will include recombinantly engineered mutants and variants, fragments, functional equivalents, derivatives, homologs and fusions of the

native selectable marker gene so long as the encoded product retains the selectable property. Useful derivatives generally have substantial sequence similarity (at the amino acid level) in regions or domains of the selectable marker associated with the selectable property. A variety of marker genes, well known to the skilled artisan, have been described, including bifunctional (i.e. positive/negative) markers (see e.g. WO 92/08796 and WO 94/28143), incorporated by reference herein. For example, selectable genes commonly used with eukaryotic cells include the genes for aminoglycoside phosphotransferase (APH), hygromycin phosphotransferase (HYG), dihydrofolate reductase (DHFR), thymidine kinase (TK), glutamine synthetase, asparagine synthetase, and genes encoding resistance to neomycin (G418), puromycin, histidinol D, bleomycin and phleomycin.

The “selectable amplifiable marker gene” usually encodes an enzyme which is required for growth of eukaryotic cells under those conditions. For example, the selectable amplifiable marker gene may encode DHFR which gene is amplified when a host cell transfected therewith is grown in the presence of the selective agent, methotrexate (MTX). Accordingly, host cells genetically modified according to any method described herein are encompassed by this invention, wherein the selectable amplifiable marker gene encodes for example for a polypeptide having the function of dihydrofolate reductase (DHFR), glutamine synthetase, CAD, adenosine deaminase, adenylate deaminase, UMP synthetase, IMP 5'-dehydrogenase, xanthine guanine phosphoribosyl transferase, HGPRTase, thymidine kinase, thymidylate synthetase, P glycoprotein 170, ribonucleotide reductase, asparagine synthetase, arginosuccinate synthetase, ornithine decarboxylase, HMG CoA reductase, acetylglucosaminyl transferase, threonyl-tRNA synthetase or Na<sup>+</sup>K<sup>+</sup>-ATPase. For a review of the exemplary selectable amplifiable marker genes see Kaufman, *Methods in Enzymology*, 185, 537-566, 1990.

One particular selectable amplifiable marker gene is the gene encoding dihydrofolate reductase (DHFR) which is necessary for the biosynthesis of purines. Cells lacking the DHFR gene will not grow on medium lacking purines. The DHFR gene is therefore useful as a dominant selectable marker to select and amplify genes in such cells growing in medium lacking purines. The selection agent used in conjunction with a DHFR gene is methotrexate (MTX).

Another selection and/or amplification marker is the glutamine synthetase (GS) gene. The GS gene encodes the glutamine synthetase enzyme which is required for synthesis of the amino acid glutamine. Cells lacking the GS gene or expressing low endogenous GS levels will not grow in glutamine-free media. The GS gene is therefore useful as a dominant selectable marker to select and amplify genes in such cells growing in glutamine-free medium. The selection agent used in conjunction with the GS gene is methionine sulfoximine (MSX).

Selection may also be made by fluorescence activated cell sorting (FACS) using for example a cell surface marker, bacterial  $\beta$ -galactosidase or fluorescent proteins (e.g. green fluorescent proteins (GFP) and their variants from *Aequorea victoria* and *Renilla reniformis* or other species; red fluorescent proteins, fluorescent proteins and their variants from non-bioluminescent species (e.g. *Discosoma* sp., *Anemonia* sp., *Clavularia* sp., *Zoanthus* sp.) to select for recombinant cells. The term “selection agent” refers to a substance that interferes with the growth or survival of a host cell that is deficient in a particular selectable gene. For example, to select for the presence of an antibiotic resistance gene like APH (aminoglycoside phosphotransferase) in a transfected cell the antibiotic

Geneticin (G418) is used. The selection agent can also comprise an “amplifying agent” which is defined for purposes herein as an agent for amplifying copies of the amplifiable gene if the selectable marker gene relied on is an amplifiable selectable marker. For example, methotrexate is a selection agent useful for the amplification of the DHFR gene.

The term “gene expression” or “expression” relates to the transcription and/or translation of a heterologous gene sequence in a host cell. The expression rate can be generally determined, either on the basis of the quantity of corresponding mRNA which is present in the host cell or on the basis of the quantity of gene product produced which is encoded by the gene of interest. The quantity of mRNA produced by transcription of a selected nucleotide sequence can be determined for example by Northern blot hybridisation, ribonuclease-RNA-protection, in situ hybridisation of cellular RNA or by PCR methods (e.g. quantitative PCR). Proteins which are encoded by a selected nucleotide sequence can also be determined by various methods such as, for example, ELISA, protein A HPLC, western blot, radioimmunoassay, immunoprecipitation, detection of the biological activity of the protein, immune staining of the protein followed by FACS analysis or fluorescence microscopy, direct detection of a fluorescent protein by FACS analysis or fluorescence microscopy or by spectrophotometry.

By “increased titer or productivity”, “increased expression” or “improved expression or productivity” is meant the increase in expression, synthesis or secretion of a heterologous sequence introduced into a host cell, for example of a gene coding for a therapeutic protein, by comparison with a suitable control, for example a protein encoded by a cDNA versus a protein encoded by an intron-containing gene. There is increased titer or productivity if a cell according to the invention is cultivated according to a method according to the invention described here, and if this cell has at least a 1.2-fold, a 1.5-fold, a two-fold, a three-fold, a four-fold or a five-fold increase in specific productivity or titer. There is also increased titer or productivity if a cell according to the invention is cultivated according to a method according to the invention described here, and if this cell has at least a 1.2-fold or at least a 1.5-fold or at least a two-fold or at least a three-fold increase in specific productivity or titer. There is also in particular increased titer or productivity if a cell according to the invention is cultivated according to a method according to the invention described here, and if this cell has at least a 1.2-fold to five-fold, preferably a 1.5-fold to five-fold, more preferably—two-fold to five-fold particularly preferably a three-fold to five-fold increase in specific productivity or titer.

An increased titer, productivity or expression may be obtained by using one of the heterologous intron-containing gene set-ups according to the invention. This may be combined with other approaches such as a FACS-assisted selection of recombinant host cells which contain, as additional selectable marker, one or more fluorescent proteins (e.g. GFP) or a cell surface marker. Other methods of obtaining increased expression, and a combination of different methods may also be used, are based for example on the use of cis-active elements for manipulating the chromatin structure (e.g. LCR, UCOE, EASE, isolators, S/MARs, STAR elements), on the use of (artificial) transcription factors, treatment of the cells with natural or synthetic agents for up-regulating endogenous or heterologous gene expression, improving the stability (half-life) of mRNA or the protein, improving the initiation of mRNA translation, increasing the gene dose by the use of episomal plasmids (based on the use of viral sequences as replication origins, e.g. SV40, polyoma, adenovirus, EBV or

BPV), the use of amplification-promoting sequences or in vitro amplification systems based on DNA concatemers.

The term “titer” is a statement of the product concentration in a defined volume, e.g. ng/mL, mg/mL, mg/L, g/L.

The term “specific productivity” refers to the amount of protein produced by the cell, in pg per cell and per day. It is calculated using the formula  $\text{pg}/((\text{Ct}-\text{Co})t/\ln(\text{Ct}-\text{Co}))$ , where Co and Ct indicate the number of cells on seeding or harvesting and t is the cultivation period.

A further embodiment of the above mentioned methods relates to a method, wherein the polypeptide(s)/product(s) which is/are encoded by the gene(s) of interest and being expressed in said host cell, is/are isolated from the cells or the cell culture supernatant, if secreted into the culture medium.

Said production cells are cultivated preferentially in serum-free medium and in suspension culture under conditions which are favorable for the expression of the desired gene(s) and isolating the protein of interest from the cells and/or the cell culture supernatant. Preferably the protein of interest is recovered from the culture medium as a secreted polypeptide, or it can be recovered from host cell lysates if expressed without a secretory signal. It is necessary to purify the protein of interest from other recombinant proteins, host cell proteins and contaminants in a way that substantially homogenous preparations of the protein of interest are obtained. As a first step often cells and/or particulate cell debris are removed from the culture medium or lysate. The product of interest thereafter is purified from contaminant soluble proteins, polypeptides and nucleic acids, for example, by fractionation on immunoaffinity or ion-exchange columns, ethanol precipitation, reverse phase HPLC, Sephadex chromatography, chromatography on silica or on a cation exchange resin such as DEAE. In general, methods teaching a skilled person how to purify a heterologous protein expressed by host cells, are well known in the art.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of cell biology, molecular biology, cell culture, immunology and the like which are in the skill of one in the art. These techniques are fully disclosed in the current literature.

The invention concerns a transcription unit or expression vector or a polynucleotide sequence comprising at least one signal peptide gene sequence operatively linked to a gene of interest, whereby said signal peptide gene sequence comprises at least one heterologous intron located within a signal peptide exon, whereby the 5' and 3' ends of said exon are defined as occurring in a corresponding native signal peptide gene sequence. In a specific embodiment of the present invention the expression of the gene of interest is at least 20% increased.

In another embodiment of the present invention the coding region of the gene of interest comprises at least one immunoglobulin domain.

In a further embodiment of the present invention said gene of interest encodes an immunoglobulin, preferably an IgG1, IgG2, or IgG4, or an Fc fusion protein, or a single chain format or another antibody-derived molecule.

In a specific embodiment of the present invention said intron is a nucleic acid sequence at least 90% identical to SEQ ID NO:1, SEQ ID NO:2 or SEQ ID NO:3.

Preferably the intron comprises/consists of/is SEQ ID NO:1, SEQ ID NO:2 or SEQ ID NO:3, more preferably the intron comprises/consists of/is SEQ ID NO:1 or SEQ ID NO:3.

The novel introns are designed to have additional functionality and/or greater efficacy than unmodified introns, e.g. they comprise sequences which can act as stop codons in case of a



29

non-splicing event of the messenger RNA leading to premature translation termination of the protein or have optimized splice donor and acceptor sequences included. Additionally, the novel introns of the present invention are especially useful when placed at a sequence position in an signal peptide exon different from the intron position(s) in the native eukaryotic gene or into a eukaryotic, prokaryotic or synthetic gene, i.e. positions which naturally do not contain an intron. When placed at such a position this leads to an at least comparable or preferably improved expression of a gene of interest, especially genes encoding antibodies and antibody derived fragments.

In another embodiment of the present invention the signal peptide sequence is derived from an immunoglobulin heavy or light chain gene.

In a specific embodiment of the present invention the signal peptide gene sequence encodes an amino acid sequence selected from the group consisting of: SEQ ID NOs: 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, and 16.

In another embodiment of the present invention the heterologous intron is located at a position resulting in a functional splice donor site and a functional splice acceptor site.

Therefore, a functional splice donor site and a functional splice acceptor site is characterized by its ability to allow splicing of the (to be) introduced intron, i.e. removal of the (heterologous) intron.

In a further embodiment of the present invention the heterologous intron is positioned within the nucleotide sequence successions CAG:C, CAG:T, AAG:C, AAG:T, TAG:T or TAG:C, preferably CAG:C, CAG:T, AAG:C or AAG:T, whereby the colons denote the site of intron insertion.

In a specific embodiment of the present invention the nucleotide sequence successions encode the amino acid pairs selected from the group consisting of:

| CAG:C  | CAG:T  | AAG:C  | AAG:T  | TAG:C  | TAG:T  |
|--------|--------|--------|--------|--------|--------|
| GlnLeu | GlnPhe | LysLeu | LysPhe | PheSer | PheSer |
| CAGCTN | CAGTTY | AAGCTN | AAGTTY | TTTAGC | TTTAGT |
| GlnPro | GlnLeu | LysPro | LysLeu | SerSer | SerSer |
| CAGCCN | CAGTTR | AAGCCN | AAGTTR | TCTAGC | TCTAGT |
| GlnHis | GlnSer | LysHis | LysSer | TyrSer | TyrSer |
| CAGCAY | CAGTCN | AAGCAY | AAGTCN | TATAGC | TATAGT |
| GlnGln | GlnTyr | LysGln | LysTyr | CysSer | CysSer |
| CAGCAR | CAGTAY | AAGCAR | AAGTAY | TGTAGC | TGTAGT |
| GlnArg | GlnCys | LysArg | LysCys | LeuSer | LeuSer |
| CAGCGN | CAGTGY | AAGCGN | AAGTGY | CTTAGC | CTTAGT |
| SerAla | GlnTrp | GlnAla | LysTrp | ProSer | ProSer |
| TCAGCN | CAGTGG | CAAGCN | AAGTGG | CCTAGC | CCTAGT |
| ProAla | SerVal | LysAla | GlnVal | HisSer | HisSer |
| CCAGCN | TCAGTN | AAAGCN | CAAGTN | CATAGC | CATAGT |
| ThrAla | ProVal | GluAla | LysVal | ArgSer | ArgSer |
| ACAGCN | CCAGTN | GAAGCN | AAAGTN | CGTAGC | CGTAGT |
| AlaAla | ThrVal | LeuSer | GluVal | IleSer | IleSer |
| GCAGCN | ACAGTN | YTAAGC | GAAGTN | ATTAGC | ATTAGT |

30

-continued

| CAG:C     | CAG:T  | AAG:C  | AAG:T  | TAG:C  | TAG:T  |
|-----------|--------|--------|--------|--------|--------|
| 5 PheSer  | AlaVal | SerSer | LeuSer | ThrSer | ThrSer |
| TTCAGC    | GCAGTN | TCAAGC | YTAAGT | ACTAGC | ACTAGT |
| SerSer    | PheSer | ProSer | SerSer | AsnSer | AsnSer |
| 10 TCCAGC | TTCAGT | CCAAGC | TCAAGT | AATAGC | AATAGT |
| TyrSer    | SerSer | GlnSer | ProSer | SerSer | SerSer |
| TACAGC    | TCCAGT | CAAAGC | CCAAGT | AGTAGC | AGTAGT |
| 15 CysSer | TyrSer | ArgSer | GlnSer | ValSer | ValSer |
| TGCAGC    | TACAGT | MGAAGC | CAAAGT | GTTAGC | GTTAGT |
| LeuSer    | CysSer | IleSer | ArgSer | AlaSer | AlaSer |
| 20 CTCAGC | TGCAGT | ATAAGC | MGAAGT | GCTAGC | GCTAGT |
| ProSer    | LeuSer | ThrSer | IleSer | AspSer | AspSer |
| CCCAGC    | CTCAGT | ACAAGC | ATAAGT | GATAGC | GATAGT |
| 25 HisSer | ProSer | LysSer | ThrSer | GlySer | GlySer |
| CACAGC    | CCCAGT | AAAAGC | ACAAGT | GGTAGC | GGTAGT |
| ArgSer    | HisSer | ValSer | LysSer | LeuAla | LeuVal |
| 30 CGCAGC | CACAGT | GTAAGC | AAAAGT | YTAGCN | YTAGTN |
| IleSer    | ArgSer | AlaSer | ValSer | IleAla | IleVal |
| ATCAGC    | CGCAGT | GCAAGC | GTAAGT | ATAGCN | ATAGTN |
| 35 ThrSer | IleSer | GluSer | AlaSer | ValAla | ValVal |
| ACCAGC    | ATCAGT | GAAAGC | GCAAGT | GTAGCN | GTAGTN |
| AsnSer    | ThrSer | GlySer | GluSer |        |        |
| 40 AACAGC | ACCAGT | GGAAGC | GAAAGT |        |        |
| SerSer    | AsnSer |        | GlySer |        |        |
| AGCAGC    | AACAGT |        | GGAAGT |        |        |
| 45 ValSer | SerSer |        |        |        |        |
| GTCAGC    | AGCAGT |        |        |        |        |
| AlaSer    | ValSer |        |        |        |        |
| 50 GCCAGC | GTCAGT |        |        |        |        |
| AspSer    | AlaSer |        |        |        |        |
| GACAGC    | GCCAGT |        |        |        |        |
| 55 GlySer | AspSer |        |        |        |        |
| GGCAGC    | GACAGT |        |        |        |        |
|           | GlySer |        |        |        |        |
| 60 GGCAGT |        |        |        |        |        |

colon: site of intron insertion, N = any base, Y = C or T, R = A or G, M = C or A

In another embodiment of the present invention said signal peptide gene sequence comprises at least one heterologous intron located within a signal peptide exon, specifically at the amino acid positions or the amino acid sequences shown for

## 31

example in FIG. 8. More specifically these amino acid sequences are encoded by the nucleotide sequences shown for example in FIG. 8.

In a specific embodiment of the present invention said transcription unit or expression vector further comprises at least one promoter and at least one terminator, optionally said transcription unit or expression vector further comprises at least one selection marker (bacterial or eukaryotic) and/or at least one enhancer. Preferably the promoter is CMV, ubiquitin, or elongation factor promoter.

Preferably the selection marker is an amplifiable selection marker like dihydrofolate reductase or glutamine synthetase and/or a non-amplifiable selection marker like neomycin, puromycin, hygromycin, or ampicillin resistance markers.

Preferably the terminator is a polyadenylation sequence from hamster growth hormone, SV40 or bovine growth hormone.

Preferably the enhancer is from CMV or SV40.

In another embodiment of the present invention said gene of interest comprises/contains at least one amino acid substitution/mutation modulating the biological (e.g. ADCC, half life, binding) and/or biophysical (e.g. stability, solubility) properties of the encoded protein. Preferably said amino acid substitution/mutation is within an immunoglobulin domain such as VH, VL, CH1 CH2, CH3, hinge region.

In a specific embodiment of the present invention the intron is a nucleic acid sequence derived from a kappa light chain intron (e.g. SEQ ID NO: 1) or derived from hamster dhfr intron (e.g. SEQ ID NO:2 or SEQ ID NO:3).

In another specific embodiment of the present invention the intron is a nucleic acid sequence at least 90% identical to SEQ ID NO:1, SEQ ID NO:2 or SEQ ID NO:3. Preferably the intron comprises/consists of/is SEQ ID NO:1, SEQ ID NO:2 or SEQ ID NO:3, more preferably the intron comprises/consists of/is SEQ ID NO:1 or SEQ ID NO:3.

The invention furthermore concerns a (host) cell comprising the transcription unit or expression vector as described above. Specifically said (host) cell is a eukaryotic cell, preferably a mammalian cell, more preferably a rodent cell, most preferably a hamster cell like Chinese hamster ovary (CHO) cell. Most preferred is a CHO DG44, or CHO DUKX (host) cell.

The invention further concerns the use of the transcription unit or the expression vector according to any one of the above embodiments of the present invention for improving the productivity of a cell.

The invention further concerns a method of producing a (heterologous/recombinant) protein of interest encoded by a polynucleotide sequence comprising at least one signal peptide gene sequence operatively linked to said polynucleotide sequence comprising the following steps:

- introducing at least one (heterologous) intron sequence into said signal peptide gene sequence by placing the intron into a nucleotide sequence of a signal peptide encoding exon, whereby the 5' and 3' ends of the exon are defined as occurring in a corresponding native signal peptide gene sequence, and
- introducing the intron-containing signal peptide gene sequence of step a) into a transcription unit or a expression vector encoding a protein of interest, and
- transfecting a cell with said transcription unit or said vector of step b), and
- cultivating said cell of step c) under conditions which allow expression of said protein of interest.

The invention further concerns a method of producing a (heterologous/recombinant) protein of interest encoded by a polynucleotide sequence comprising at least one signal pep-

## 32

tide gene sequence operatively linked to said polynucleotide sequence comprising the following steps:

- introducing said polynucleotide sequence comprising at least one signal peptide gene sequence into a transcription unit or a (mammalian) expression vector, and
- introducing at least one heterologous intron sequence into said signal peptide gene sequence by placing the intron into a nucleotide sequence of a signal peptide encoding exon, whereby the 5' and 3' ends of the exon are defined as occurring in a corresponding native signal peptide gene sequence, and
- transfecting a cell with said transcription unit or said vector of step b), and
- cultivating said cell of step c) under conditions which allow expression of said protein of interest.

The invention furthermore concerns a method of producing a (heterologous/recombinant) protein of interest encoded by a polynucleotide sequence comprising at least one domain of an immunoglobulin gene comprising:

- transfecting a cell with the transcription unit or expression vector according to the invention,
- cultivating said cell of step a) under conditions which allow expression of said protein of interest.

In a specific embodiment of any one of the described methods according to the present invention the intron sequence is at least 90% identical to SEQ ID NO:1, SEQ ID NO:2 or SEQ ID NO:3. Preferably the intron comprises/consists of/is SEQ ID NO:1, SEQ ID NO:2 or SEQ ID NO:3, more preferably the intron comprises/consists of/is SEQ ID NO:1 or SEQ ID NO:3.

In another embodiment of any one of the described methods according to the present invention such method additionally comprises the (following) step of isolating and purifying said protein of interest.

The invention further concerns a method of generating a recombinant host cell/production cell comprising:

- transfecting a cell with the transcription unit or expression vector according to the present invention, and
- selecting a recombinant host cell/production cell.

The invention furthermore concerns a kit comprising a vector including at least parts of a transcription unit comprising a first nucleotide sequence comprising at least one (heterologous) intron located within a signal peptide encoding exon, whereby the 5' and 3' ends of said exon are defined as occurring in a corresponding native signal peptide gene sequence, and whereby the nucleotide sequence of this transcription unit comprises suitable recognition sites for restriction enzymes allowing the integration or operative linkage of a second nucleotide sequence coding for a gene of interest (preferably an immunoglobulin, such as an IgG1, IgG2, or IgG4, or an Fc fusion protein, or a single chain format or another antibody-derived molecule), and instructions and optionally a vector map,

optionally a host cell, optionally a cultivation medium for the cultivation of a host cell, and/or

optionally a selection medium for selecting and cultivating a transfected host cell.

Effectively the 2 nucleotide sequences in step a) are/become operatively linked.

The following examples are not limiting. They merely show possible embodiments of the invention. A person skilled in the art could easily adjust the conditions to apply it to other embodiments.

## EXPERIMENTAL

## Abbreviations

AP: Alkaline phosphatase  
 ch: Chimeric  
 CH1: CH1 domain of immunoglobulin heavy chain  
 CH2: CH2 domain of immunoglobulin heavy chain  
 CH3: CH3 domain of immunoglobulin heavy chain  
 CHO: Chinese hamster ovary  
 CL: CL domain of IgG light chain  
 DHFR: Dihydrofolate reductase  
 ELISA: Enzyme-linked immunosorbent assay  
 h: Human  
 H: Hinge region of immunoglobulin heavy chain  
 HGH: Hamster growth hormone  
 HT: Hypoxanthine/thymidine  
 HRPO: Horseradish peroxidase  
 hu: Humanized  
 IgG: Immunoglobulin G  
 mAb: Monoclonal antibody  
 NPT: Neomycin phosphotransferase  
 PCR: Polymerase chain reaction  
 SEAP: Secreted alkaline phosphatase  
 VH: Variable region of IgG heavy chain  
 VL: Variable region of IgG light chain  
 Materials and Methods  
 Expression Vectors

Eukaryotic expression vectors are derivatives of the pAD-CMV1 vector (WO 9201055) and mediate constitutive expression of the heterologous genes driven by the CMV promoter/enhancer. For termination and polyadenylation of the transcript of the gene of interest vectors contain the polyadenylation signal of the hamster growth hormone (SEQ ID NO: 8 in EP2009059399). The pBI-26 vector encodes a hamster-derived DHFR mini gene as amplifiable selection marker (see for example EP 0 393 438) whereas the pBI-49 vector encodes an attenuated NPT gene as selection marker (SEQ ID NO: 17 and SEQ ID NO: 18 in WO 2004/050884) under the control of the SV40 early promoter and a thymidine kinase polyadenylation signal (FIG. 1).

Genes of interest encoding for heavy (IgG1, IgG2 or IgG4 isotype) and light chain (kappa or lambda) of chimeric, humanized or human monoclonal antibodies or Fc fusion proteins are cloned into the vectors using the multiple cloning sites located between promoter and polyadenylation signal. All recombinant expression vectors used as controls in the invention are summarized in FIG. 2. The genes are either cloned as cDNA or as genomic DNA versions, whereby the heavy chains and the Fc fusion proteins are cloned into the vector pBI-26 and the light chains into the vector pBI-49. In the latter the intron and exon sequences of the constant domains are derived from the natural genomic gene sequences. The introns are in their natural positions located always within the codon encoding for the first amino acid of a constant domain or the hinge region. Sequences of the variable regions of both heavy and light chain of all antibodies are de novo synthesized and optimized to remove potential cryptic splice sites, direct repeats, secondary structure elements and other motifs interfering with expression. Furthermore the GC content for improved RNA stability and the codon usage are optimized. The gene structures of the various genes of interests are shown schematically in FIG. 3.

## Cell Culture

CHO-DG44/dhfr<sup>-/-</sup> cells are maintained in suspension culture in the serum-free medium CHO—S—SFMII (Invitrogen) supplemented with hypoxanthine and thymidine

(HT). Cells are incubated in cell culture flasks at 37° C. in a humidified atmosphere containing 5% CO<sub>2</sub>. The cell number as well as the cell viability are determined with a Cedex (Innovatis AG, Germany) or via trypan blue dye exclusion. Cells are seeded at a concentration of 1-3×10<sup>5</sup> cells/mL in fresh medium every two to three days.

## Transfections

Transfections of CHO-DG44 cells are conducted using Lipofectamine Plus reagent (Invitrogen). Per transfection 6×10<sup>5</sup> exponentially growing cells in 0.8 mL hypoxanthine/thymidine (HT)-supplemented CHO—S—SFMII medium (Invitrogen) are seeded in a well of a 6-well chamber. A mixture of plasmid DNA, 4 µL Lipofectamine and 6 µL Plus reagent in a volume of 200 µL is generated for each transfection and added to the cells, following the protocol of the manufacturer. After incubation for 3 hours 2 mL of HT-supplemented CHO—S—SFMII medium is added. Transient transfections are performed in triplicate for each vector combination and supernatants are harvested 2 to 3 days post transfection.

## ELISA

mAb titers are quantified by ELISA with standard protocols using a goat anti-human IgG Fc fragment antibody (Jackson ImmunoResearch Europe Ltd., UK) and an AP-conjugated goat anti-human kappa light chain antibody or AP-conjugated goat anti-human lambda light chain antibody (Sigma). Titers of Fc fusion proteins are quantified by ELISA with standard protocols using a mouse anti-human IgG (Fc specific) antibody (Sigma) and an HRP-conjugated goat anti-human IgG (Fc specific) antibody (Sigma). For detection the fast p-nitrophenyl phosphate tablet set (Sigma) or the TMB ELISA substrate (Serva) are used and the reaction is stopped either with NaOH (AP substrates) or H<sub>2</sub>SO<sub>4</sub> (HRP substrates). Purified mAb antibody of the same isotype as the expressed mAb or Fc fusion protein is used as standard. Samples are analyzed using an Infinite M200 Pro reader (TECAN, Crailsheim, Germany). Adsorptions are measured either at 405 nm (AP substrates) or 450 nm (HRP substrates).

## SEAP Assay

SEAP activity is determined with the Great EscAPE SEAP Chemiluminescence Assay according to the protocol of the manufacturer (Clontech). Samples are analyzed using an Infinite M200 Pro reader (TECAN, Crailsheim, Germany).

## EXAMPLES

## Example 1

## Cloning of Heterologous Intron-Containing Immunoglobulin G and Fc Fusion Genes

The first intron sequence (SEQ ID NO:1, FIG. 4A) is based on the intron sequence located between the variable and constant region of a human kappa gene. It is modified to introduce a single BglII restriction site close to the 5' end of the intron for cloning purposes introduce sequences which can act as stop codons in case of a non-splicing event of the messenger RNA and would lead to premature translation termination of the protein introduce a conserved branch site in the 3' region of the intron to allow for more efficient splicing.

The intron sequence is synthesized de novo at Invitrogen using the GENEART technology.

For placement of the modified kappa intron within the immunoglobulin signal peptide sequence in a position other than the natural position within the codon for the amino acid at position -4 (counting backwards from the 3' end of the

35

amino acid sequence of the signal peptide) the cloning is done as follows. The intron sequence is amplified by PCR introducing via the PCR primers at the 5' ends of the intron sequence signal peptide coding sequences and a HindIII restriction site and at the 3' ends a SgrAI restriction site. Alternatively, signal peptide coding sequences and an EcoRV restriction site are introduced at the 3' ends of the amplified fragment. In the following, the amplified fragments are cloned directionally into the HindIII and SgrAI or HindIII and EcoRV digested expression vectors encoding the cDNA sequences of the various antibodies heavy and light chain genes or Fc fusion protein genes listed in FIG. 2 (suffix "c") and shown schematically in FIG. 3 thereby exchanging the signal peptide sequences without an intron with signal sequences with a heterologous intron. The resulting recombinant vectors encoding heavy chains, light chains or Fc fusion proteins are marked with the suffix "g1" (see FIG. 5). The intron is placed in such a way into the coding region of the signal peptide that a consensus splice donor and splice acceptor sequence is generated using the appropriate codons for the amino acid threonine at position -7 and alanine at position -6 (FIG. 6).

The second intron sequence is isolated from the hamster dihydrofolate reductase gene (SEQ ID NO: 2, FIG. 4B). It is amplified from the eukaryotic expression vector pBI-26. This vector encodes a hamster-derived DHFR mini gene as amplifiable selection marker (see for example EP 0 393 438).

For placement of the dhfr intron within the immunoglobulin signal peptide sequence in a position other than the natural position within the codon for the amino acid at position -4 (counting backwards from the 3' end of the amino acid sequence of the signal peptide) the cloning is done as follows. The intron sequence is amplified by PCR introducing via the PCR primers at the 5' ends of the intron sequence signal peptide coding sequences and a HindIII restriction site and at the 3' ends a SgrAI restriction site. Alternatively, signal peptide coding sequences and an EcoRV restriction site are introduced at the 3' ends of the amplified fragment. In the following, the amplified fragments are cloned directionally into the HindIII and SgrAI or HindIII and EcoRV digested expression vectors encoding the cDNA sequences of the heavy and light chain genes or Fc fusion protein genes listed in FIG. 2 (suffix "c") and shown schematically in FIG. 3 thereby exchanging the signal peptide sequences without an intron with signal sequences with a heterologous intron. The resulting recombinant vectors encoding heavy chains, light chains or Fc fusion proteins are marked with the suffix "g4" (see FIG. 5). The intron is placed in such a way into the coding region of the signal peptide that an efficient splice donor and splice acceptor sequence is generated using the appropriate codons for the amino acid threonine at position -7 and alanine at position -6 (FIG. 6).

The third intron sequence (SEQ ID NO:3) is based on the intron sequence isolated from the hamster dihydrofolate reductase gene (SEQ ID NO: 2) but is further modified to

introduce a single BglII restriction site close to the 5' end of the intron for cloning purposes

introduce optimized splice donor and acceptor sequences.

The intron sequence is synthesized de novo at Invitrogen using the GENEART technology.

For placement of the modified dhfr intron within the immunoglobulin signal peptide sequence in a position other than the natural position within the codon for the amino acid at position -4 (counting backwards from the 3' end of the amino acid sequence of the signal peptide) the same approach as described for the non-modified hamster dhfr intron (SEQ ID NO:2) is used. The resulting recombinant vectors encoding

36

genes for heavy or light chain or an Fc fusion protein and containing the heterologous intron within the signal peptide sequence are marked with the suffix "g6" (see FIG. 5). The intron is placed in such a way into the coding region of the signal peptide that an efficient splice donor and splice acceptor sequence is generated using the appropriate codons for the amino acid threonine at position -7 and alanine at position -6 (FIG. 6).

### Example 2

#### Impact of Heterologous Introns on Expression of Immunoglobulin G1

To evaluate the impact of the intron sequences derived from human kappa gene (SEQ ID NO: 1) and the hamster dihydrofolate reductase (SEQ ID NO: 2) on the expression if placed within the exon regions of IgG1 molecules transient transfections are performed. Transient transfections allow assessment of the expression independent of chromosomal integration sites. The introns are placed within the immunoglobulin signal peptide sequence of heavy and light chain (see FIGS. 5 and 6). CHO-DG44 cells are co-transfected with vectors encoding the heavy and light chain of an antibody ( $6.5 \times 10^{10}$  molecules per vector). This set-up is tested with two different mouse/human chimeric antibodies (=chIgG1 and chIgG1B), a humanized antibody (=hIgG1) and a human antibody (=hIgG1). The latter contains a lambda light chain, all other antibodies contain kappa light chains. As a control CHO-DG44 cells are co-transfected with vector combinations encoding either the cDNA of the respective antibodies or the genomic DNA with introns positioned as in the natural genes within the codon encoding the first amino acid of the constant domains or hinge region (see FIGS. 2 and 3). Aside of the different antibody gene formats the genetic set-ups of the various vectors for the expression of the various antibodies are identical.

Supernatants are harvested 2 to 3 days post transfection and the IgG1 titers are determined using ELISA. Per vector combination 3 cell pools are transfected in each transfection series. At least two independent transfection series for each set of antibodies are performed. To correct for transfection efficiency cells are co-transfected with the plasmid pCMV-SEAP (100 ng DNA/transfection reaction), which encodes the secreted alkaline phosphatase, and the SEAP activity is measured.

FIG. 7 shows the data of at least 2 independent transient transfection series performed in triplicate. Overall, cells transfected with heterologous intron containing genes show 1.7-3.8 fold higher expression of the various antibodies compared to cells transfected with vectors encoding the cDNAs of the respective antibodies even though the sequences of the variable regions are optimized. Surprisingly, all heterologous intron-containing gene set-ups in which a modified single intron sequence derived from a human kappa gene (SEQ ID NO: 1, FIG. 4A) is placed in a new position within the signal peptide sequence of heavy and light chain genes are outperforming the corresponding natural genomic gene versions. The natural intron derived from the hamster dihydrofolate reductase gene (SEQ ID NO: 2, FIG. 4B) leads as single intron to an expression comparable to the corresponding natural genomic gene version with 3 introns if placed within the signal peptide sequence of the heavy chain and the kappa light chain.

37

## Example 3

## Impact of Heterologous Introns on Expression of Immunoglobulin G2

To evaluate the impact of the intron sequences derived from human kappa gene (SEQ ID NO: 1) on the expression if placed within the exon regions of an IgG2 molecule transient transfections are performed. Transient transfections allow assessment of the expression independent of chromosomal integration sites. The intron is placed within the signal peptide sequence of the heavy chain and the kappa light chain (see FIGS. 5 and 6). CHO-DG44 cells are co-transfected with vectors encoding the heavy and light chain of an antibody ( $6.5 \times 10^{10}$  molecules per vector). This set-up is tested with a mouse/human chimeric antibody (=chIgG2B) which contains a kappa light chain. As a control CHO-DG44 cells are co-transfected with vector combinations encoding either the cDNA of the respective antibody or the genomic DNA with introns which are positioned as in the natural genes within the codon encoding the first amino acid of the constant domains or hinge region (see FIGS. 2 and 3). Aside of the different antibody gene formats the genetic set-ups of the various vectors for the expression of the various antibodies are identical.

Supernatants are harvested 2 to 3 days post transfection and the IgG2 titers are determined using ELISA. Per vector combination 3 cell pools are transfected in each transfection series. Two independent transfection series for each set of antibodies are performed. To correct for transfection efficiency cells are co-transfected with the plasmid pCMV-SEAP (100 ng DNA/transfection reaction), which encodes the secreted alkaline phosphatase, and the SEAP activity is measured.

FIG. 7 shows the data of 2 independent transient transfection series performed in triplicate. Overall, cells transfected with heterologous intron containing genes show 2.4-3 fold higher expression of the mouse/human chimeric IgG2 molecule compared to cells transfected with vectors encoding the cDNA of this antibody even though the sequences of the variable regions are optimized. Surprisingly, the heterologous intron-containing gene set-ups in which a modified single intron sequence derived from a human kappa gene (SEQ ID NO: 1, FIG. 4A) is placed in a new position within the signal peptide sequence of heavy and light chain genes are outperforming the corresponding natural genomic gene versions.

## Example 4

## Impact of Heterologous Introns on Expression of Immunoglobulin G4

To evaluate the impact of the intron sequences derived from human kappa gene (SEQ ID NO: 1) on the expression if placed within the exon regions of an IgG4 molecule transient transfections are performed. Transient transfections allow assessment of the expression independent of chromosomal integration sites. The intron is placed within the signal peptide sequence of the heavy chain and of the kappa light chain (see FIGS. 5 and 6). CHO-DG44 cells are co-transfected with vectors encoding the heavy and light chain of an antibody ( $6.5 \times 10^{10}$  molecules per vector). This set-up is tested with a

38

mouse/human chimeric antibody (=chIgG4B) which contains a kappa light chain. As a control CHO-DG44 cells are co-transfected with vector combinations encoding either the cDNA of the respective antibody or the genomic DNA with introns which are positioned as in the natural genes within the codon encoding the first amino acid of the constant domains or hinge region (see FIGS. 2 and 3). Aside of the different antibody gene formats the genetic set-ups of the various vectors for the expression of the various antibodies are identical.

Supernatants are harvested 2 to 3 days post transfection and the IgG4 titers are determined using ELISA. Per vector combination 3 cell pools are transfected in each transfection series. Two independent transfection series for each set of antibodies are performed. To correct for transfection efficiency cells are co-transfected with the plasmid pCMV-SEAP (100 ng DNA/transfection reaction), which encodes the secreted alkaline phosphatase, and the SEAP activity is measured.

FIG. 7 shows the data of 2 independent transient transfection series performed in triplicate. Overall, cells transfected with heterologous intron containing genes show 2.5-5.3 fold higher expression of the mouse/human chimeric IgG4 molecule compared to cells transfected with vectors encoding the cDNAs of this antibody even though the sequences of the variable regions are optimized. Surprisingly, the heterologous intron-containing gene set-ups in which a modified single intron sequence derived from a human kappa gene (SEQ ID NO: 1, FIG. 4A) is placed in a new position within the signal peptide sequence of heavy and light chain genes are outperforming the corresponding natural genomic gene versions.

## Example 5

## Impact of Heterologous Introns on Expression of Fc Fusion Protein

To evaluate the impact of the intron sequences derived from the human kappa gene (SEQ ID NO: 1) or from the hamster dhfr gene (SEQ ID NO:2 or SEQ ID NO:3) on the expression if placed within the signal peptide sequence of an Fc fusion protein (FIGS. 5 and 6) transient transfections are performed. Transient transfections allow assessment of the expression independent of chromosomal integration sites. CHO-DG44 cells are transfected with vectors encoding the various Fc fusion proteins ( $13 \times 10^{10}$  molecules). This set-up is tested with a fusion protein consisting of the wild type Fc sequence of a human IgG1 or a Fc mutant with a Leu234Ala and Leu235Ala (according to EU numbering) substitution in the CH2 domain. As a control CHO-DG44 cells are transfected with vector combinations encoding either the cDNA of the respective Fc fusion protein or the genomic DNA with a natural intron within the codon for the first amino acid of the CH3 domains (see FIGS. 2 and 3). Aside of the different gene formats the genetic set-ups of the various vectors for the expression of the various Fc fusion proteins are identical.

Supernatants are harvested 2 to 3 days post transfection and the titers of the Fc fusion proteins are determined using ELISA. Per vector combination 3 cell pools are transfected in each transfection series. Two independent transfection series for each set of Fc fusion proteins are performed. To correct for transfection efficiency cells are co-transfected with the plasmid pCMV-SEAP (100 ng DNA/transfection reaction), which encodes the secreted alkaline phosphatase, and the SEAP activity is measured.

## SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 28

<210> SEQ ID NO 1

<211> LENGTH: 360

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic intron polynucleotide derivative derived from genomic region of human kappa light chain and further modified

<400> SEQUENCE: 1

```
gtaagtgcac tttcctaata gatctaattc taaactctga gggggctcgga tgacgtggcc    60
attctttgcc taaagcattg agtttactgc aaggtcagaa aagcatgcaa agccctcaga    120
atggctgcaa agagctccaa caaaacaatt tagaacttta ttaaggaata gggggaagct    180
aggaagaaac tcaaacatc aagattttaa atacgcttct tggctctcctt gctataatta    240
tctgggataa gcattgctgt tctgtctgt ccctaacatg ccctgtgatt atccgcaaac    300
aacacaccca agggcagaac ttgtttatac taacaccatc ctgtttgctt ctttcctcag    360
```

<210> SEQ ID NO 2

<211> LENGTH: 304

<212> TYPE: DNA

<213> ORGANISM: Cricetus griseus

<220> FEATURE:

<223> OTHER INFORMATION: Intron sequence derived from genomic region of hamster dihydrofolate reductase

<400> SEQUENCE: 2

```
gtactggctg gattgggtta gggaaaccga ggcggttcgc tgaatcgggt cgagcacttg    60
gcgagagacg cggggccaac tacttaggga cagtcattgag gggtaggccc gccggctgct    120
gcccttgccc atgcccggg tgatcccat gctgtgccag cctttgccc gaggcgtctt    180
agctgggagc aaagtccggt cactgggcag caccaccccc cggaacttga tgggtagccg    240
ctgagatgga gccctgagc acgtgacagg gtccctgtta acgcagtgtt tctctaactt    300
tcag                                              304
```

<210> SEQ ID NO 3

<211> LENGTH: 308

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic intron polynucleotide derived from genomic region of hamster dihydrofolate reductase and further modified

<400> SEQUENCE: 3

```
gtaagtctg gattgggtta gatctggaaa ccgaggcggt tcgctgaatc gggctcagca    60
cttgccggag acgcgcgggc caactactta gggacagtca tgaggggtag gcccgccggc    120
tgctgccctt gccatgccc gcggtgatcc ccatgctgtg ccagcctttg ccagaggcg    180
ctctagctgg gagcaaagtc cggtcactgg gcagaccac cccccgact tgcattggta    240
gccgctgaga tggagcctga gcacacgtga cagggtccct gttaacgcag tgtttctctc    300
cctttcag                                              308
```

<210> SEQ ID NO 4

<211> LENGTH: 19

<212> TYPE: PRT

<213> ORGANISM: Mus musculus

-continued

---

<220> FEATURE:  
<223> OTHER INFORMATION: Signal peptide of mouse heavy chain

<400> SEQUENCE: 4

Met Gly Trp Ser Cys Ile Ile Leu Phe Leu Val Ala Thr Ala Thr Gly  
1            5                    10                    15

Val His Ser

<210> SEQ ID NO 5  
<211> LENGTH: 20  
<212> TYPE: PRT  
<213> ORGANISM: Mus musculus  
<220> FEATURE:  
<223> OTHER INFORMATION: Signal peptide of mouse kappa light chain

<400> SEQUENCE: 5

Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro  
1            5                    10                    15

Gly Ser Ser Gly  
            20

<210> SEQ ID NO 6  
<211> LENGTH: 19  
<212> TYPE: PRT  
<213> ORGANISM: Mus musculus  
<220> FEATURE:  
<223> OTHER INFORMATION: Signal peptide of mouse heavy chain

<400> SEQUENCE: 6

Met Gly Trp Ser Trp Ile Phe Leu Phe Leu Leu Ser Gly Thr Ala Gly  
1            5                    10                    15

Val Leu Ser

<210> SEQ ID NO 7  
<211> LENGTH: 19  
<212> TYPE: PRT  
<213> ORGANISM: Homo sapiens  
<220> FEATURE:  
<223> OTHER INFORMATION: Signal peptide of human heavy chain

<400> SEQUENCE: 7

Met Asp Trp Thr Trp Arg Val Phe Cys Leu Leu Ala Val Ala Pro Gly  
1            5                    10                    15

Ala His Ser

<210> SEQ ID NO 8  
<211> LENGTH: 22  
<212> TYPE: PRT  
<213> ORGANISM: Homo sapiens  
<220> FEATURE:  
<223> OTHER INFORMATION: Signal peptide of human kappa light chain

<400> SEQUENCE: 8

Met Asp Met Arg Val Pro Ala Gln Leu Leu Gly Leu Leu Leu Leu Trp  
1            5                    10                    15

Phe Pro Gly Ala Arg Cys  
            20

<210> SEQ ID NO 9  
<211> LENGTH: 22  
<212> TYPE: PRT  
<213> ORGANISM: Homo sapiens  
<220> FEATURE:  
<223> OTHER INFORMATION: Signal peptide of human kappa light chain

-continued

---

<400> SEQUENCE: 9

Met Asp Met Arg Val Pro Ala Gln Leu Leu Gly Leu Leu Leu Trp  
1                   5                   10                   15  
  
Leu Pro Gly Ala Lys Cys  
                  20

<210> SEQ ID NO 10  
<211> LENGTH: 19  
<212> TYPE: PRT  
<213> ORGANISM: Homo sapiens  
<220> FEATURE:  
<223> OTHER INFORMATION: Signal peptide of human heavy chain

<400> SEQUENCE: 10

Met Glu Phe Gly Leu Ser Trp Leu Phe Leu Val Ala Ile Leu Lys Gly  
1                   5                   10                   15  
  
Val Gln Cys

<210> SEQ ID NO 11  
<211> LENGTH: 20  
<212> TYPE: PRT  
<213> ORGANISM: Homo sapiens  
<220> FEATURE:  
<223> OTHER INFORMATION: Signal peptide of human kappa light chain

<400> SEQUENCE: 11

Met Glu Ala Pro Ala Gln Leu Leu Phe Leu Leu Leu Trp Leu Pro  
1                   5                   10                   15  
  
Asp Thr Thr Gly  
                  20

<210> SEQ ID NO 12  
<211> LENGTH: 19  
<212> TYPE: PRT  
<213> ORGANISM: Mus musculus  
<220> FEATURE:  
<223> OTHER INFORMATION: Signal peptide of mouse heavy chain

<400> SEQUENCE: 12

Met Ala Val Leu Gly Leu Leu Phe Cys Leu Val Thr Phe Pro Ser Cys  
1                   5                   10                   15  
  
Val Leu Ser

<210> SEQ ID NO 13  
<211> LENGTH: 20  
<212> TYPE: PRT  
<213> ORGANISM: Mus musculus  
<220> FEATURE:  
<223> OTHER INFORMATION: Signal peptide of mouse kappa light chain

<400> SEQUENCE: 13

Met Val Ser Thr Pro Gln Phe Leu Val Phe Leu Leu Phe Trp Ile Pro  
1                   5                   10                   15  
  
Ala Ser Arg Ser  
                  20

<210> SEQ ID NO 14  
<211> LENGTH: 20  
<212> TYPE: PRT  
<213> ORGANISM: Mus musculus  
<220> FEATURE:  
<223> OTHER INFORMATION: Signal peptide of mouse kappa light chain

<400> SEQUENCE: 14



-continued

---

Met Glu Thr Asp Thr Ile Leu Leu Trp Val Leu Leu Leu Trp Val Pro  
1 5 10 15

Gly Ser Thr Gly  
20

<210> SEQ ID NO 15  
<211> LENGTH: 19  
<212> TYPE: PRT  
<213> ORGANISM: Mus musculus  
<220> FEATURE:  
<223> OTHER INFORMATION: Signal peptide of mouse heavy chain

<400> SEQUENCE: 15

Met Asn Phe Gly Ser Ser Leu Ile Phe Leu Val Leu Val Leu Lys Gly  
1 5 10 15

Val Gln Cys

<210> SEQ ID NO 16  
<211> LENGTH: 19  
<212> TYPE: PRT  
<213> ORGANISM: Mus musculus  
<220> FEATURE:  
<223> OTHER INFORMATION: Signal peptide of mouse heavy chain

<400> SEQUENCE: 16

Met Gly Trp Ser Cys Ile Phe Leu Phe Leu Val Ala Thr Ala Thr Gly  
1 5 10 15

Val His Ser

<210> SEQ ID NO 17  
<211> LENGTH: 27  
<212> TYPE: PRT  
<213> ORGANISM: Homo sapiens  
<220> FEATURE:  
<223> OTHER INFORMATION: Signal peptide of human intercellular adhesion molecule

<400> SEQUENCE: 17

Met Ala Pro Ser Ser Pro Arg Pro Ala Leu Pro Ala Leu Leu Val Leu  
1 5 10 15

Leu Gly Ala Leu Phe Pro Gly Pro Gly Asn Ala  
20 25

<210> SEQ ID NO 18  
<211> LENGTH: 23  
<212> TYPE: PRT  
<213> ORGANISM: Homo sapiens  
<220> FEATURE:  
<223> OTHER INFORMATION: Signal peptide of human monocyte chemoattractant protein 1

<400> SEQUENCE: 18

Met Lys Val Ser Ala Ala Leu Leu Cys Leu Leu Leu Ile Ala Ala Thr  
1 5 10 15

Phe Ile Pro Gln Gly Leu Ala  
20

<210> SEQ ID NO 19  
<211> LENGTH: 26  
<212> TYPE: PRT  
<213> ORGANISM: Homo sapiens  
<220> FEATURE:  
<223> OTHER INFORMATION: Signal peptide of human oncostatin M

<400> SEQUENCE: 19

-continued

---

Met Gly Val Leu Leu Thr Gln Arg Thr Leu Leu Ser Leu Val Leu Ala  
1 5 10 15

Leu Leu Phe Pro Ser Met Ala Ser Met Ala  
20 25

<210> SEQ ID NO 20  
<211> LENGTH: 24  
<212> TYPE: PRT  
<213> ORGANISM: Homo sapiens  
<220> FEATURE:  
<223> OTHER INFORMATION: Signal peptide of human acid alpha glucosidase

<400> SEQUENCE: 20

Met Gly Val Arg His Pro Pro Cys Ser His Arg Leu Leu Ala Val Cys  
1 5 10 15

Ala Leu Val Ser Leu Ala Thr Ala  
20

<210> SEQ ID NO 21  
<211> LENGTH: 20  
<212> TYPE: PRT  
<213> ORGANISM: Homo sapiens  
<220> FEATURE:  
<223> OTHER INFORMATION: Signal peptide of human Fc gamma receptor III

<400> SEQUENCE: 21

Met Trp Gln Leu Leu Leu Pro Thr Ala Leu Leu Leu Val Ser Ala  
1 5 10 15

Gly Met Arg Thr  
20

<210> SEQ ID NO 22  
<211> LENGTH: 22  
<212> TYPE: PRT  
<213> ORGANISM: Homo sapiens  
<220> FEATURE:  
<223> OTHER INFORMATION: Signal peptide of human placental alkaline phosphatase

<400> SEQUENCE: 22

Met Leu Gly Pro Cys Met Leu Leu Leu Leu Leu Leu Gly Leu Arg  
1 5 10 15

Leu Gln Leu Ser Leu Gly  
20

<210> SEQ ID NO 23  
<211> LENGTH: 22  
<212> TYPE: PRT  
<213> ORGANISM: Homo sapiens  
<220> FEATURE:  
<223> OTHER INFORMATION: Signal peptide of human tumor necrosis factor receptor

<400> SEQUENCE: 23

Met Ala Pro Val Ala Val Trp Ala Ala Leu Ala Val Gly Leu Glu Leu  
1 5 10 15

Trp Ala Ala Ala His Ala  
20

<210> SEQ ID NO 24  
<211> LENGTH: 7  
<212> TYPE: PRT  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic

-continued

---

peptide

<400> SEQUENCE: 24

Val Ala Thr Ala Thr Gly Val  
1 5

<210> SEQ ID NO 25  
<211> LENGTH: 10  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> SEQUENCE: 25

gtggccacag 10

<210> SEQ ID NO 26  
<211> LENGTH: 11  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> SEQUENCE: 26

ccaccggcgt g 11

<210> SEQ ID NO 27  
<211> LENGTH: 45  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
oligonucleotide  
<220> FEATURE:  
<221> NAME/KEY: misc\_feature  
<222> LOCATION: (1)..(40)  
<223> OTHER INFORMATION: This region may encompass 11-40 nucleotides,  
wherein some positions may be absent  
<220> FEATURE:  
<221> NAME/KEY: modified\_base  
<222> LOCATION: (41)..(41)  
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 27

YYYYYYYYYY YYYYYYYYYY YYYYYYYYYY YYYYYYYYYY nyagr 45

<210> SEQ ID NO 28  
<211> LENGTH: 200  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
polynucleotide  
<220> FEATURE:  
<221> NAME/KEY: misc\_feature  
<222> LOCATION: (1)..(200)  
<223> OTHER INFORMATION: This region may encompass 100-200 nucleotides,  
wherein some positions may be absent

<400> SEQUENCE: 28

aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 60

aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 120

aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 180

aaaaaaaaa aaaaaaaaaa 200

---

## 51

The invention claimed is:

1. A transcription unit or expression vector comprising at least one signal peptide gene sequence operatively linked to a gene of interest, whereby said signal peptide gene sequence comprises at least one heterologous intron located within a signal peptide exon, whereby the 5' and 3' ends of said exon are defined as occurring in a corresponding native signal peptide gene sequence.

2. The transcription unit or expression vector of claim 1, wherein the coding region of the gene of interest comprises at least one immunoglobulin domain.

3. The transcription unit or expression vector of claim 1, whereby said gene of interest encodes an immunoglobulin.

4. The transcription unit or expression vector of claim 1, wherein said heterologous intron is a nucleic acid sequence at least 90% identical to SEQ ID NO:1, SEQ ID NO:2 or SEQ ID NO:3.

5. The transcription unit or expression vector of claim 1, wherein the signal peptide sequence is derived from an immunoglobulin heavy or light chain gene.

6. The transcription unit or expression vector of claim 1, whereby the signal peptide gene sequence encodes an amino acid sequence selected from the group consisting of: SEQ ID NOs: 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, and 16.

7. The transcription unit or expression vector of claim 1, wherein the heterologous intron is located at a position resulting in a functional splice donor site and a functional splice acceptor site.

8. The transcription unit or expression vector of claim 1, wherein the heterologous intron is positioned within the nucleotide sequence successions CAG:C, CAG:T, AAG:C, AAG:T, TAG:T or TAG:C, preferably CAG:C, CAG:T, AAG:C or AAG:T, whereby the colons denote the site of intron insertion.

9. The transcription unit or expression vector of claim 8, wherein the nucleotide sequence successions encode the amino acid pairs selected from the group consisting of:

| CAG:C  | CAG:T  | AAG:C  | AAG:T  | TAG:C  | TAG:T  |
|--------|--------|--------|--------|--------|--------|
| GlnLeu | GlnPhe | LysLeu | LysPhe | PheSer | PheSer |
| CAGCTN | CAGTTY | AAGCTN | AAGTTY | TTTAGC | TTTAGT |
| GlnPro | GlnLeu | LysPro | LysLeu | SerSer | SerSer |
| CAGCCN | CAGTTR | AAGCCN | AAGTTR | TCTAGC | TCTAGT |
| GlnHis | GlnSer | LysHis | LysSer | TyrSer | TyrSer |
| CAGCAY | CAGTCN | AAGCAY | AAGTCN | TATAGC | TATAGT |
| GlnGln | GlnTyr | LysGln | LysTyr | CysSer | CysSer |
| CAGCAR | CAGTAY | AAGCAR | AAGTAY | TGTAGC | TGTAGT |
| GlnArg | GlnCys | LysArg | LysCys | LeuSer | LeuSer |
| CAGCGN | CAGTGY | AAGCGN | AAGTGY | CTTAGC | CTTAGT |
| SerAla | GlnTrp | GlnAla | LysTrp | ProSer | ProSer |
| TCAGCN | CAGTGG | CAAGCN | AAGTGG | CCTAGC | CCTAGT |
| ProAla | SerVal | LysAla | GlnVal | HisSer | HisSer |
| CCAGCN | TCAGTN | AAAGCN | CAAGTN | CATAGC | CATAGT |
| ThrAla | ProVal | GluAla | LysVal | ArgSer | ArgSer |
| ACAGCN | CCAGTN | GAAGCN | AAAGTN | CGTAGC | CGTAGT |

## 52

-continued

| CAG:C  | CAG:T  | AAG:C  | AAG:T  | TAG:C  | TAG:T  |
|--------|--------|--------|--------|--------|--------|
| AlaAla | ThrVal | LeuSer | GluVal | IleSer | IleSer |
| GCAGCN | ACAGTN | YTAAGC | GAAGTN | ATTAGC | ATTAGT |
| PheSer | AlaVal | SerSer | LeuSer | ThrSer | ThrSer |
| TTCAGC | GCAGTN | TCAAGC | YTAAGT | ACTAGC | ACTAGT |
| SerSer | PheSer | ProSer | SerSer | AsnSer | AsnSer |
| TCCAGC | TTCAGT | CCAAGC | TCAAGT | AATAGC | AATAGT |
| TyrSer | SerSer | GlnSer | ProSer | SerSer | SerSer |
| TACAGC | TCCAGT | CAAAGC | CCAAGT | AGTAGC | AGTAGT |
| CysSer | TyrSer | ArgSer | GlnSer | ValSer | ValSer |
| TGCAGC | TACAGT | MGAAGC | CAAAGT | GTTAGC | GTTAGT |
| LeuSer | CysSer | IleSer | ArgSer | AlaSer | AlaSer |
| CTCAGC | TGCAGT | ATAAGC | MGAAGT | GCTAGC | GCTAGT |
| ProSer | LeuSer | ThrSer | IleSer | AspSer | AspSer |
| CCCAGC | CTCAGT | ACAAGC | ATAAGT | GATAGC | GATAGT |
| HisSer | ProSer | LysSer | ThrSer | GlySer | GlySer |
| CACAGC | CCCAGT | AAAAGC | ACAAGT | GGTAGC | GGTAGT |
| ArgSer | HisSer | ValSer | LysSer | LeuAla | LeuVal |
| CGCAGC | CACAGT | GTAAGC | AAAAGT | YTAGCN | YTAGTN |
| IleSer | ArgSer | AlaSer | ValSer | IleAla | IleVal |
| ATCAGC | CGCAGT | GCAAGC | GTAAGT | ATAGCN | ATAGTN |
| ThrSer | IleSer | GluSer | AlaSer | ValAla | ValVal |
| ACCAGC | ATCAGT | GAAAGC | GCAAGT | GTAGCN | GTAGTN |
| AsnSer | ThrSer | GlySer | GluSer |        |        |
| AACAGC | ACCAGT | GGAAGC | GAAAGT |        |        |
| SerSer | AsnSer |        | GlySer |        |        |
| AGCAGC | AACAGT |        | GGAAGT |        |        |
| ValSer | SerSer |        |        |        |        |
| GTCAGC | AGCAGT |        |        |        |        |
| AlaSer | ValSer |        |        |        |        |
| GCCAGC | GTCAGT |        |        |        |        |
| AspSer | AlaSer |        |        |        |        |
| GACAGC | GCCAGT |        |        |        |        |
| GlySer | AspSer |        |        |        |        |
| GGCAGC | GACAGT |        |        |        |        |
|        | GlySer |        |        |        |        |
|        | GGCAGT |        |        |        |        |

colon: site of intron insertion, N = any base, Y = C or T, R = A or G, M = C or A

10. The transcription unit or expression vector according to claim 1, wherein said transcription unit or expression vector further comprises at least one promoter and at least one ter-

53

minator, optionally said transcription unit or expression vector further comprises at least one selection marker and/or at least one enhancer.

11. A host cell comprising the transcription unit or expression vector according to claim 1.

12. The host cell of claim 11, wherein said cell is a eukaryotic cell.

13. A method of producing a protein of interest encoded by a polynucleotide sequence comprising at least one signal peptide gene sequence operatively linked to said polynucleotide sequence comprising the following steps:

- a) introducing at least one heterologous intron sequence into said signal peptide gene sequence by placing the intron into a nucleotide sequence of a signal peptide encoding exon, whereby the 5' and 3' ends of the exon are defined as occurring in a corresponding native signal peptide gene sequence, and
- b) introducing the heterologous intron-containing signal peptide gene sequence of step a) into a transcription unit or a expression vector encoding a protein of interest, and
- c) transfecting a cell with said transcription unit or said vector of step b), and
- d) cultivating said cell of step c) under conditions which allow expression of said protein of interest.

14. A method of producing a protein of interest encoded by a polynucleotide sequence comprising at least one domain of an immunoglobulin gene comprising:

- a) transfecting a cell with the transcription unit or expression vector of claim 1,
- b) cultivating said cell of step a) under conditions which allow expression of said protein of interest.

54

15. The method according to claim 13, whereby the heterologous intron sequence is at least 90% identical to SEQ ID NO:1, SEQ ID NO:2 or SEQ ID NO:3.

16. The method of claim 13, additionally comprising the following step of isolating and purifying said protein of interest.

17. A method of generating a recombinant host cell/production cell comprising:

- a) transfecting a cell with the transcription unit or expression vector of claim 1, and
- b) selecting a recombinant host cell/production cell.

18. A kit comprising

- a) a vector including at least parts of a transcription unit comprising a first nucleotide sequence comprising at least one heterologous intron located within a signal peptide encoding exon, whereby the 5' and 3' ends of said exon are defined as occurring in a corresponding native signal peptide gene sequence, and whereby the nucleotide sequence of this transcription unit comprises suitable recognition sites for restriction enzymes allowing the integration or operative linkage of a second nucleotide sequence coding for a gene of interest, and instructions and optionally a vector map, and optionally a host cell, and optionally a cultivation medium for the cultivation of a host cell, and/or optionally a selection medium for selecting and cultivating a transfected host cell.

19. The transcription unit or expression vector of claim 3, whereby said gene of interest encodes an IgG1, IgG2, or IgG4, or an Fc fusion protein, or a single chain format or another antibody-derived molecule.

\* \* \* \* \*